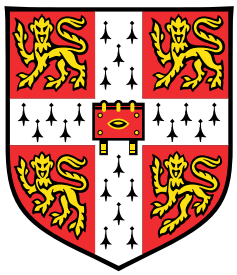
Assessing Health Vulnerability to Air Pollution in Seoul Using an Agent-Based Simulation



Hyesop Shin

Department of Geography
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

Fitzwilliam College

October 2020

To my beloving two J's, Jireh and Jane ...

Declaration

I hereby declare that this dissertation is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University of similar institution except as declared in the Preface and specified in the text. This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. I confirm this dissertation contains fewer than 80,000 words, in line with the regulations of the University of Cambridge Department of Geography Degree Committee.

Hyesop Shin
October 2020

Assessing Health Vulnerability to Air Pollution in Seoul using an Agent-based Simulation

Hyesop Shin

Abstract

This study aims to investigate the exposure to air pollution and the consequent health effects in Seoul, South Korea, and suggest possible solutions using agent-based modelling (ABM). ABM is a useful technique that can simulate pollution generation and exposure, mobility patterns of unique individuals, and explore future scenarios. The first study compared Universal Kriging and Generalised Additive Models to spatially interpolate pollution station data over Seoul. A new method was discovered to enhance the accuracy of the NO_2 prediction on roads. Next, ABM was used to evaluate potential health loss for a set of demographic groups after being cumulatively exposed to particulates (PM_{10}), with a nominal heath impact threshold of $100\mu g/m^3$. Finally, a traffic simulation examined the coupled problem of non-exhaust emissions and behaviour, and estimated exposure to PM_{10} for groups of drivers and pedestrians in central Seoul. Having tested the sensitivity to calibrated parameters, scenarios of traffic restriction and modification of pedestrian behaviour to avoid polluted areas was investigated. With less difference between interpolation methods, the result showed a remarkable contrast between roadside and background NO_2 as well as a daily cycle, while PM_{10} had a small variance over the time of the day but had greater seasonal oscillation. The first ABM study showed that disparities in health may arise as a result of differences in socioeconomic status, especially when a group was exposed over a long period, and road proximity caused additional health loss. In the traffic simulation study, extreme PM_{10} was found along roadways, but although drivers were exposed to extreme values, longer exposure for pedestrians led to higher health risks. Despite the absence of reliable data linking exposure to actual health effects, it is possible to make progress with ABM. In addition, pollution exposure can vary by commuting patterns and the urban development of one's location. Scenario forecasting such as these can be advantageous for healthcare policy - to aid the most vulnerable groups and districts.

Keywords: Air pollution, Health vulnerability, Agent-based model, Seoul

Acknowledgements

I begin this acknowledgement by praising the Heavenly Father for his guidance through the long journey. His words have lifted me whenever I was down and weary, hovering here and there. I believe that, through the gospel, my faith grew stronger and the way of setting my whole life became firm in Him.

I also express my deep gratitude to my supervisor, Mike Bithell, for his meticulous support on my writing, experimental design, and even the know-how to code when I struggled. His warm comments and encouraging critiques have nourished my knowledge foundation, to think critically and clearly. This will lead me to become a better scholar in the future.

I truly thank my buddies in the department - Oliver, Peadar, JP, Annemarie, and Debolina. Thanks for offering the time to hang out whenever we(I) felt bored, wanted some good food under the sun, or needed some help/encouragement. I can't remember how much cups of coffee we drank per week! If it weren't for you guys, my PhD life might have been really boring. I deeply send my gratitude to Matt for helping me out with my English writing and speaking during my first year in the UK. My English improved a lot thanks to you! I also thank my close fellows in the Korean Church including Pastor Hong, Prof. Justin Jeon, Rev.Hankyu Kim, Yeonsik, Minkyu, Choong-il, James, and Bohwan. I will never forget the times we spent together at church, in the freezing caravans, and out in the fields discussing topics overnight, from divinity to family life and vocational goals. Thanks for telling me how to love God and love others.

This thesis wouldn't have been possible without the support of my family from Korea. Despite being far away, I thank my parents and my parents-in-law for the endless support and sacrifice to end this journey successfully. I am really proud that you are my parents and you really deserve more than words. Also, thanks for sending

parcels whenever we crave food and Korean items. We remember jumping up and down when we received the parcel and the heartwarming letters.

Lastly, I would like to show my sincere appreciation and love to my wife and daughter, Jireh and Jane, who have seen the beginning and the end of my PhD. I learnt that the most difficult job on earth is being a PhD student's wife, but you have done it really well! My lovely daughter Jane, thank you for coming to this world and bringing a great joy to us. I promise to show you the bright future.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

1.1 Research rationale

The adverse health effects of pollutants are an important concern for our society. Industry, power plants, vehicles, and domestic appliances emit pollutants that lead to smog, haze, and acidification in the urban atmosphere (Guarnieri and Balmes, 2014). The effect of some pollutants is extremely dangerous because pollutants are generally invisible and can easily penetrate and accumulate in the human body, regardless of the type of substance, either particulates or gases. Their toxicity may eventually lead to various ailments, such as cardiovascular or pulmonary diseases (Guarnieri and Balmes, 2014; Wardekker et al., 2012).

In recent decades, megacities in East Asia have been facing a number of health challenges driven by air pollution (see Figure 1.1). Beijing, for instance, had an extremely high amount of $PM_{2.5}$ concentrations at around 534µg/m^3 on February 26, 2014, which was far beyond the moderate level of 100µg/m^3 (Liu et al., 2016). Seoul is another pollution-risk city where the city's PM_{10} level remained at 50µg/m^3 during 2016, but has lately experienced several pollution episodes, of which the most severe level was 162µg/m^3 on March 5th-6th 2019 (Lee and Park, 2019). This exceeds the levels of the World Health Organisation guidelines for PM_{10} of 20µg/m^3 as an annual mean, and 50µg/m^3 as a 24-hour mean (WHO, 2012). However, Asian megacities, despite their rapid growth during recent decades, have yet to set stringent regulations for emission control. Although the recent announcement by China to aim to be carbon neutral by 2060, still some giant cities, including Beijing, and Seoul, have already agreed

to add more coal-powered plants (as coal is perceived to be the most cost-effective substance for energy production). As a consequence, these citizens are more likely to be exposed to pollutants generated from local industrial sites, as well as from vehicle emissions, and atmospheric import from other areas (Wardekker et al., 2012; WHO, 2012)

The situation in urban areas is particularly problematic because the exposure level can vary by socioeconomic status and individual behaviour. Evidence has shown that people who live close to roads or industrial sites are socially deprived, and that the young and the old are more likely to be vulnerable to pollution (Beevers et al., 2013; David and Don, 2012; Guarnieri and Balmes, 2014; Heinrich et al., 2013; Li et al., 2012; O'Neill et al., 2003); in addition, drivers are exposed to much higher risk compared to other transport users due to the ingress of particulates and emission exhausts through filters and recirculation (Amato, 2018; Kreider et al., 2019). These problems, therefore, have highlighted the need to discover vulnerabilities to pollution across populations, and to develop tools to help prevent future health problems.

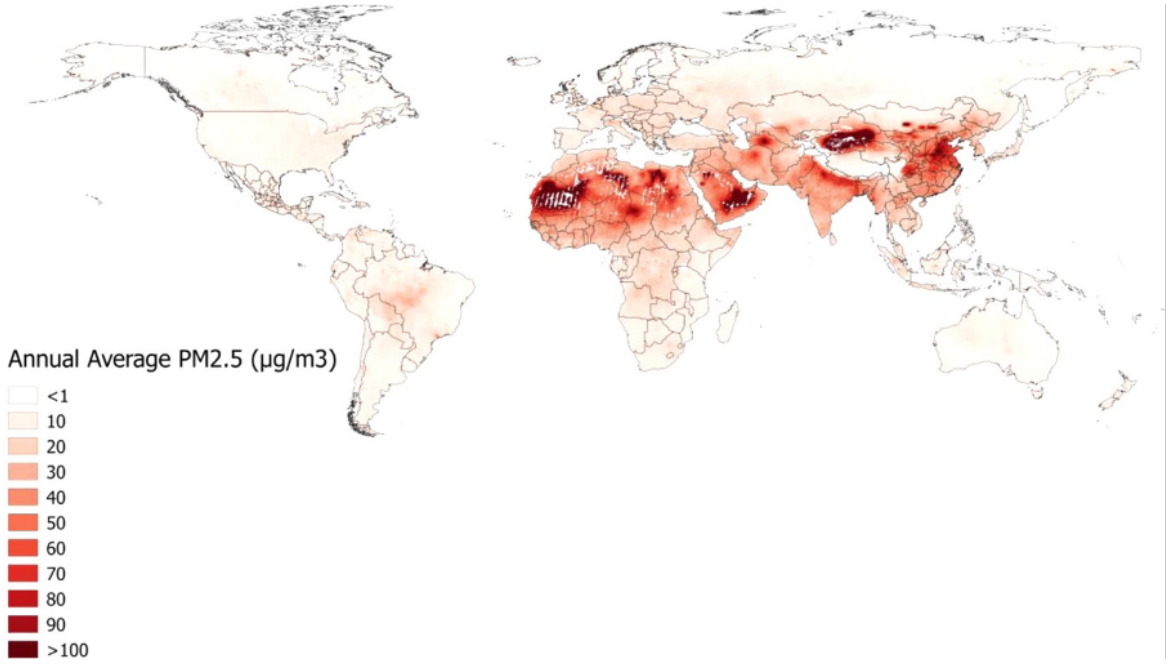


Fig. 1.1 Estimated 2013 annual average of $PM_{2.5}$ (particulates less than 2.5 micrometres in diameter) in the troposphere as detected by the grid cell average of satellite-based estimates from the TM5-FASST simulation. (imported from Brauer et al. (2016))

In an effort to help assessments of air pollution and health problems in various regions, there have been attempts at developing the concept of health vulnerability. According to Parry (2007), vulnerability is defined as "the susceptibility of individuals to the impacts of hazards as well as their resilience to recover from them", and this is made up of three components: *exposure*, as in the amount of influence from air conditions; *sensitivity*, as in the individual's biophysical susceptibility and characteristics associated with social status; and *adaptive capacity*, as in the ability or resilience that a population can adapt to harmful air conditions (Parry, 2007). Renaud (2006) has explained vulnerability measurement as a means of identifying (locating) places in need of assistance, as well as a tool for monitoring progress and analysing trends.

1.2 Key Criticisms from Previous Studies

Despite the growing importance of measuring health exposure and vulnerability, weaknesses exist in terms of how the concepts and ideas outlined in the previous paragraph apply in the case of exposure in air pollution. Some of the gaps in the published literature include the following points where:

First, predicting air quality maps with 'temporally rich but spatially poor' resolution monitors will not fully account for the variability in space. Air pollution is highly variable in terms of space and time, which ranges from minutes to years, and from centimetres to kilometres. However, previous exposure studies that used spatial interpolation or dispersion models may have used sparsely located monitors to produce maps that were visually smoothed. Many exposure studies have used IDW (Inverse Distance Weighted), Kriging (Ordinary, Block, Co-), Generalised additive models to collect the information from disciplines researching pollution (Dewulf et al., 2016; Greater London Authority, 2017; Min et al., 2020; Nyhan et al., 2016; Son et al., 2010; Wong et al., 2004). Although the overall output seems visually meaningful, the predicted maps were too smooth to the point that it neglected the small-scale variability that usually happens between pollution monitoring stations. In reality, pollution levels vary by building blocks and even by the height of buildings (Aristodemou et al., 2018). These studies also used seasonally or annually averaged values to understand the aggregated levels of PM_{10}, which smoothed away the extremes of pollution levels. Hence, ignoring granular spatial and temporal scale pollution will cause ecological fallacies when evaluating population exposure.

Second, earlier studies have assessed the exposure to air pollution without considering individual behaviours and susceptibility. Previous works have often assumed that residents have equal exposure to air pollutants which is far from the case in practice (Atkinson et al., 2016; Halonen et al., 2016; Heinrich et al., 2013; Moreno-Jiménez et al., 2016; Samoli et al., 2016). Exposure can vary in individuals depending not only on their demographic groups, socio-economic status (i.e. where they go), and residential addresses (i.e. where they live) but also on their physical condition and activity patterns (i.e. where they go). Some studies that apply OD (Origin-Destination) matrices have estimated exposure patterns based on people's home and workplaces; however, the matrices only give starting and ending points that lack the trajectory the individuals have taken (Dewulf et al., 2016; Nyhan et al., 2016; Xu et al., 2019). Mobile phone data collected from cell towers provide positional information within a boundary, however, this can only display the location of a biased population who use a mobile service (Xu et al., 2019). Because people move across space, more accurate measurements of activity patterns and pollution concentrations need to be achieved (Hwang and Lee, 2018).

Third, studies have neglected the effort to project future circumstances that can avoid further adverse health outcomes. Many epidemiological studies have attempted to understand exposure and air pollution through historic data. The factors that represented an individual or population were put into a model, analysed and were interpreted. Retrospective research is useful because of the depth of the data and the tools which facilitate interpretation. However, they lack the capabilities that are needed to understand what might happen in the future (Hammond, 2015). Prospective scenarios, such as climate change or air pollution change, anticipate a variety of possible outcomes with possible uncertainties, but this allows people to understand where different interventions can lead in the long term (Hammond, 2015; Wilensky and Rand, 2015).

Most of these limitations come from a deterministic and linear approach in which the assumptions smooth spatial features of pollution emissions and people's movements. Therefore, there is a need to consider social phenomena using a bottom-up (individual-based) approach that also accounts for the dynamics of exposure.

1.3 Research Motivation

Agent-based modelling (ABM) has been spotlighted as a novel approach for simulating actions and interactions between a collection of discrete entities and can capture aspects of emergence as a result of interactions (Bonabeau, 2002; Gilbert, 2008). Compared to statistical studies, ABM enables heterogeneous agents to move and interact across different locations without the need to aggregate up to population level, and to allow learning behaviour so that the structure of the feedback loops themselves can change over time. The dynamics of integrating agents and space have become significant in exposure modelling not only because it combines agents and space in a multi-scalar frame, but also it takes into account different representations of air quality during the agent's movement over time (Gilbert, 2008; Stanilov, 2012).

Rather than relying on static monitoring stations to derive pollution data, ABMs can also provide the pollution levels at an exact location so that there is less mismatch for assessing exposure levels. One way of doing this is to input pollution sources, vehicle emission or factory chimneys, for example, to understand how pollution is generated over space and diluted in time, and how individuals are exposed according to their activity patterns. This way of modelling will hopefully lead to better understanding the causal relationship between pollution and disease. A simple example of the approach is the conceptual model called the "Urban Suite Model – Pollution", part of the Netlogo modelling environment, that examines the fragile equilibrium of a predator-prey system where predators are pollution patches and prey are the population (Felsen and Wilensky, 2007).

Lastly, ABM can be used as a testing ground to experiment hypotheses and to apply scenarios for future decisions, in other words, "what if..." scenarios (Stanilov, 2012). The models are not used as tools to predict what will happen but rather used to explain the behaviour of the system (Epstein, 2008; Hammond, 2015; Stanilov, 2012). For this thesis, scenario forecasting is especially important because clinical trials of air pollution exposure are difficult to conduct. Numerous experiments can be conducted in the *in silico* environment.

1.4 Research Questions and Objectives

Based on the motivation of this research and the gaps in the existing literature, the research questions and objectives are described as follows.

Overall Aim

The overall purpose of this study is to estimate how health vulnerability from an unequal distribution of individuals can be affected by air pollution in the case of Seoul, South Korea. These results will provide new insights for suggesting a reduction policy for the population's future exposure.

Key Questions

This thesis addresses the following questions that fall under the research aim:

1. How can vulnerability to air pollution be conceptualised to take in the full account of exposure and sensitivity?
2. What do spatial and temporal trends of pollution reveal about its distribution and how can the methods be used to estimate pollution in a large urban space?
3. How does the response pattern from pollution vary between population groups and how does it appear on the geographical landscape after applying an agent-based model?
4. What do GIS and agent-based modelling scenarios suggest for the case-study city about future patterns of risk and the need for interventions, e.g. in terms of protection and support?

Research Objectives

1. To construct a conceptual framework for exposure assessment by reviewing the existing literature about pollution patterns, human exposure, and actual disease symptoms
2. To understand the spatial and temporal trends of pollutants in Seoul, and interpolation and visualisation techniques

3. To select and test a variety of exposure-response models to estimate how exposure might translate into disease, and validate patient data
4. To investigate the health, social, and environmental impacts of pollution episodes in both short and long-term periods using an agent-based model
5. To discover the changes in population groups or areas at risk in different scenarios

The novelty of this work is that it uses open data and methodology to tackle the research questions. The data and codes are stored in a public repository[1].

1.5 Thesis Structure

The remainder of this thesis is presented in 5 chapters (see Table 1.1). Chapter 2 investigates the existing literature to better elucidate the process of pollution patterns, human exposure, and actual disease symptoms (*Research Objective 1*). The chapter then harmonises the relationship between pollution-exposure, exposure-activity, and exposure-health outcomes to understand the sensitivity of adverse health effects, analysed by demographic groups and socioeconomic status in many case studies. As this study focuses on the movement of individuals through ABM, the chapter reviews simulation studies that take into account activity patterns and behavioural responses in relation to air pollution.

Using 57 stations around greater Seoul, Chapter 3 explores a time-series of PM_{10} and NO_2 to reveal the temporal distribution of pollutants in various aspects (*Research Objective 2*). Following this, the chapter compares two spatial interpolation methods, General additive models (GAM) and Universal Kriging (UK), to estimate how the pollutants vary over the urban scale. GAM is one of the methods used to statistically infer unknown values between observations based on unknown smooth functions of indicators. UK is also a statistical method but predicts the unknown values based on the weights of the average of the observations in the neighbourhood areas, which are governed by the semivariogram (Kumar, 2007). As both methods have weaknesses in predicting small scale variances, this study also considered adding weights to the road networks. The results present two points during the day in summer 2013 and winter

[1]Chapter 3 in https://www.github.com/mrsensible/SpInterpolation, Chapter 4 in https://www.github.com/mrsensible/PollutionABM, and Chapter 5 in https://www.github.com/mrsensible/SeoultrafficABM.

2014. The chapter ends with a discussion of how spatial estimations can be used to account for health effects and the pros and cons of using spatial interpolation methods in pollution studies.

Chapter 4 evaluates the health effects of long-term exposure to PM_{10} in Seoul districts. This chapter begins with the idea that the response from a cumulative exposure can vary by which demographic group one belongs and where one moves to (*Research Objective 4*). This chapter applies point-to-point movement to infer the population's activity patterns, with a variation by demographic subgroups. Even though the modelled PM_{10} results from Chapter 3 have the same temporal scale of 12 hours with this chapter, the chapter did not import the results because the interpolated pollution fields are made out of the mean and the variance, which may miss out the extremes of PM_{10}. A health loss parameter was developed to associate health outcomes with PM_{10} exposure with a nominal threshold of $100\mu g/m^3$. The sensitivity of the parameter was tested and calibrated against hospital admitted patients with respiratory disorders (*Research Objective 3*). The study applies scenarios of long-term pollution change and adjustments of health resilience across all 25 districts (*Research Objective 5*).

Chapter 5 investigates how traffic simulations can be used to model emission and estimated exposure to PM_{10} by different groups of commuters (*Research Objective 4,5*). The chapter simulates the movement of pedestrians and vehicles on a minute-by-minute basis and generates non-exhaust emissions that come from tyre and brake wear. The sensitivity of the rate of inflow cars, the vehicle emissions and dilution patterns, health loss (with the same parameters as the previous chapter), and walking speed were tested. PM_{10} at a sample roadside patch was calibrated against the roadside PM_{10}. After testing, the model executed 'what if' scenarios of vehicle restrictions and pathfinding algorithms (*Research Objective 5*).

Chapter 6 concludes by discussing the contribution of the thesis of exposure assessment using an agent-based framework and its implications for future health research. This study explores the topics in five areas: (i) data biases and missingness, (ii) issues related to exposure-response, (iii) scenarios for future health applications, and (iv) the benefits of aiming for an open computing environment.

Table 1.1 Thesis Structure by Research Objectives and Chapters

Research Objective	Chapter
1. To construct a conceptual framework for exposure assessment by reviewing the existing literature surrounding pollution patterns, human exposure, and actual disease symptoms	Chapter 2
2. To understand the spatial and temporal trends of pollutants in Seoul, and interpolation and visualisation techniques	Chapter 3
3. To select and test a variety of exposure-response models to estimate how exposure might translate into disease and validate patient data	Chapter 4
4. To investigate the health, social, and environmental impacts of pollution episodes in a short or long-term period using an agent-based model	Chapter 4, Chapter 5
5. To discover the changes in population groups or areas at risk in different scenarios	Chapter 4, Chapter 5
General Discussion	Chapter 6

1.6 An Overview of Seoul

1.6.1 Geography

Seoul was selected as a case study for this project. Seoul is the capital of South Korea and is located at 37°33'57"N, and 126°58'42"E. The total area of Seoul is 605.21 km^2 (233.67 sq. mi), and this accounts for 0.6% of the whole country. The area of Seoul known as *Tukbyulshi* (translated as 'special metropolitan area') had a population of 9,904,312 as per the 2015 census. The city has a population density of 16,365/km^2, with approximately 3.9 million households. Furthermore, there are over 8 million commuting to the city every day (Kwon et al., 2015).

Seoul was not only selected due to its high local air pollution levels and its proximity to China (from which there might be an advection of polluted air) (Bae et al., 2020), but also because the city landscape surrounded by high peaks causes air circulation difficulties. As can be seen in Figure 1.2, the city is surrounded by 26 peaks, with Mt.Bukhan (836m), in the northern highlands, Mt.Gwanak (632m) in the south, and Mt.Acha (350m) to the east. There are also hilly terrains dotted across the city, which also disturb wind flows.

In general, the typical climate in Seoul is warm and mild throughout the year, as the annual mean temperature indicates 10-15°C. However, there are tremendous temperature variations from -6 to 3°C in January to 23-26°C in August, which, in other words, means it is humid during summer and dry and windy during winter. This is because the geographical location is heavily affected by two seasonal winds: the Siberian winds during winter that bring both mild and freezing air masses, and the Pacific winds during summer that have the effect of contributing hot, humid, and muggy air (Shin, 2016).

Annual precipitation in Seoul is reported to be between 1400mm and 1500mm, but over 60% of it falls during the monsoon period: June, July, and August (Shin, 2016). Average wind speed in a year is 2.3m/s, but the velocity peaks up to 2.8m/s in spring and dips to 1.9m/s in autumn. With the topography and climatological features, it is inferred that the worst pollution in Seoul is likely to occur during the transition between winter and spring as well as autumn and winter.

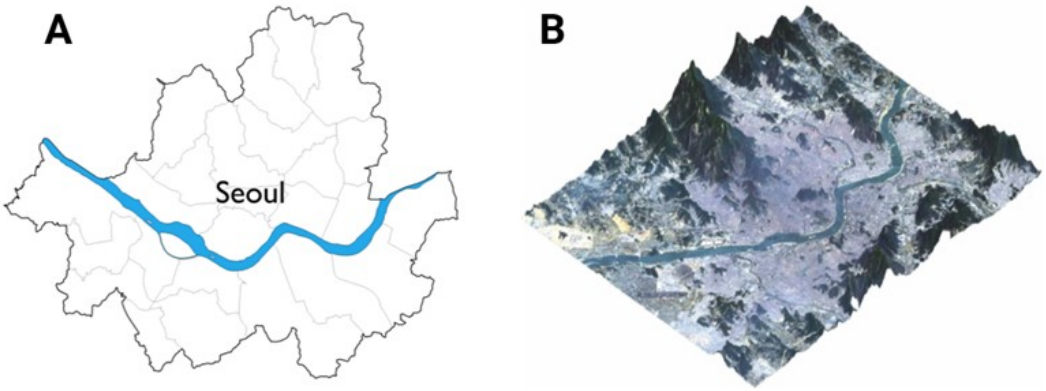


Fig. 1.2 Basemap (A) and Topography (B) of Seoul

1.6.2 Pollution Emission in Seoul

The National Environmental Research Institute (NIER), affiliated to the Ministry of Environment, established a special law for air quality improvement in 2005, aiming to reduce the pollution levels to the level stated in the WHO recommendation. After the data release was legalised in 2013, the institute published the pollution emissions of Seoul and adjacent cities on their new website (http://airemiss.nier.go.kr/).

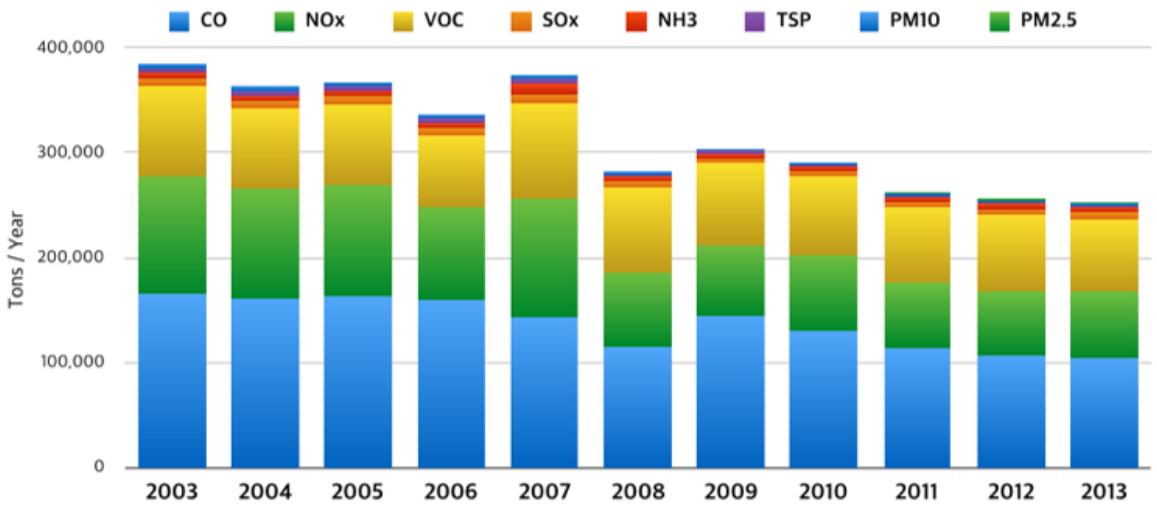


Fig. 1.3 Annual Amount of Pollution Emission in Seoul (Units: Tonnes/Year(s)). For the reports and statistics, please refer to http://airemiss.nier.go.kr/

Figure 1.3 illustrates a brief summary of pollution emissions in Seoul in 2003-2013. Eight categories of pollutants include: carbon monoxide (CO), nitrogen dioxide (NO_x), sulphur dioxide (SO_x), total suspended particles (TSP), particulates in 10 and 2.5 micro-diameters (PM_{10}, $PM_{2.5}$), volatile organic compounds (VOCs), and ammonia (NH_3). Carbon monoxide was Seoul's largest pollutant throughout the entire period, but it declined by 38%, from 166,006 to 103,039 tonnes, between 2003-2013. VOCs and NO_x are the next most abundant pollutants, which account for just over half the level of CO. Slightly more than 1,500 tonnes per year of $PM_{2.5}$ has been emitted after the institute commenced its measurement. Even though Seoul's annual emissions have declined remarkably compared to the past, the table still shows more than 250,000 tonnes of emissions are reported in Seoul, of which 180,000 tonnes of particles and 70,000 tonnes of gaseous substances may cause respiratory disorders and disorders of cardiovascular metabolism.

Figure 1.4 describes air pollution emissions of Seoul in 2013 by various sources[2]. Amongst them, road mobile pollution sources and non-road mobile pollution sources account for 56.7% of the overall emissions. This implies that more than half of the emissions are generated from transport. The transport sources additionally produce a

[2]Road mobile pollution sources: vehicle emission gases (diesel, gasoline), Non-road mobile pollution sources: heavy industries (bulldozers, excavators), ships, agricultural machines Units: Tonnes/Year

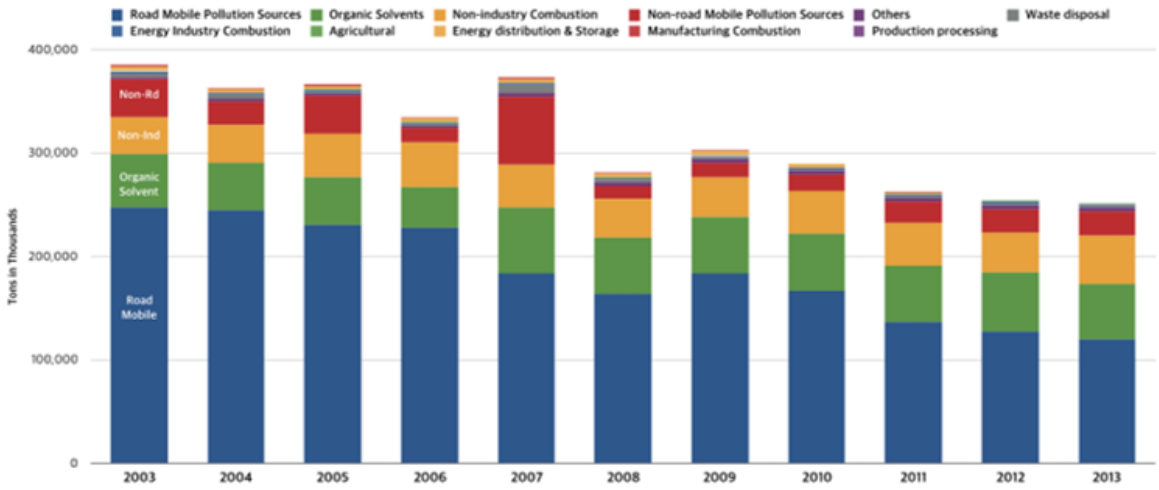


Fig. 1.4 Seoul Air Pollution Emissions by Sources. For the reports and statistics, please refer to http://airemiss.nier.go.kr/

great proportion of carbon monoxide, nitrogen dioxide, and volatile organic compounds (VOCs) at 40.8%, 25.4%, and 27.6% respectively.

With regard to $PM_{2.5}$, the amount of emissions in Seoul in 2013 was about 1.5 thousand tonnes, which is only 2% of the national amount of 76 kilo-tonnes. Over all sources, both road and non-road mobile pollution sources were the main pollutants that accounted for 677 tonnes (44.2%) and 598 tonnes (39.1%) each. Since documents less frequently report emissions at the city scale, this could only be compared between countries. From published working papers and journals, it was found that the $PM_{2.5}$ emission in the UK reached 70,000 tonnes in 2009 (Air Quality Expert Group, 2012), which was slightly less than that of Korea, while in 2010, Germany and China were the countries that emitted more than Korea at 100,000 and 638,000 tonnes respectively (Guan et al., 2014).

1.6.3 Car Ownership

According to the previous section, the use of roads has been a primary factor in emissions. Thus, this section particularly focuses on the frequency of car ownership by years as well as fuel types.

Table 1.2 presents car ownership in Seoul by vehicle registration and personal vehicles in between 2010-2016. It is clear that both total and personal vehicles have

Table 1.2 Car Ownership in Seoul by years (Unit: millions)

Type	2010	2011	2012	2013	2014	2015	2016
Personal Vehicles	2.28	2.30	2.32	2.34	2.39	2.43	2.47
Total Vehicles	2.98	2.97	2.97	2.97	3.01	3.06	3.08
%	(76.6)	(77.3)	(78)	(78.6)	(79.2)	(79.7)	(80.3)

gradually increased during the period. A total of 2.98 million vehicles were registered in 2010, and registrations then rose continuously to 3.08 million vehicles, which increased the total number of vehicles by 275,000. As for personal vehicles, the figure was 2.28 million in 2010, and it increased, similarly to total vehicles, to 2.47 million in 2016, meaning a difference of approximately 90,000 vehicles. It is clear that personal vehicle owners account for most of the total registrations in the current decade.

The total number of vehicles by fuel type during 2016 in Seoul is shown in Figure 1.5. Gasoline is indicated as the most common type by 1,596,819 vehicles, followed by diesel engines at around 1.1 million vehicles. LPG vehicles were in third place, which the figures not just above 10%. In terms of vehicle types, the proportion of cars was more than half, and gasoline engines were significantly greater in number than any other vehicles. Other vehicle types, by contrast, mostly used diesel engines, such as vans (74.7%), lorries (88.3%), and special vehicles (e.g. fork-lift, cranes) (97.1%). To summarise, in 2016, gasoline vehicles were the most common out of the total number of vehicles in Seoul, but diesel engines were the most numerous in other vehicle types. Thus, it is presumed that tailpipe emissions would have been significantly generated from these vehicles, and the compositions of the total vehicle emissions will be applied to the traffic model.

Electric vehicles have been registered in recent years, and they made up around around 1% in 2016 (although it is not illustrated in the table). Seoul is anticipating reducing the amount of exhaust emissions substantially, targeting below the legal limit of the WHO. However, Timmers and Achten (2018) insisted that even if all vehicles were electric, there would still be substantial particulate emissions from brake and tyre wear Kreider et al. (2019), which has received less attention than tail-pipe emissions.

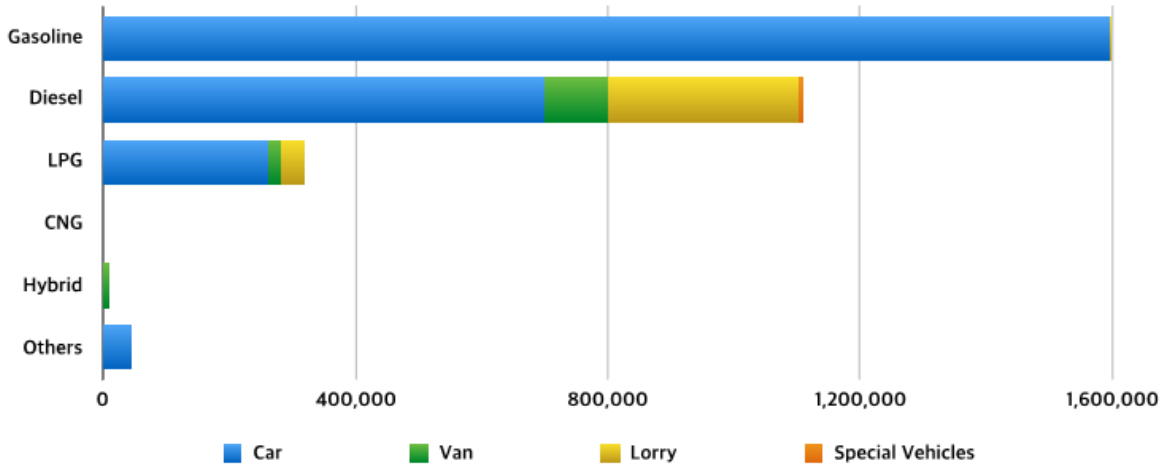


Fig. 1.5 Total Vehicles in Seoul 2016 by fuel types

1.6.4 Migration Patterns

The population of Seoul and its surrounding area, the *Sudogwon*, have been continuously increasing as the national economy has developed, and also as the 'baby boomers' reproduced to create a new generation. The rate of population growth in the period 1975-2015 is shown in Figure 1.6. The yellow bars account for the overall population in South Korea every 5 years. The blue and green line shows the rate of the population in the Sudogwon and the non-Sudogwon respectively. It clearly shows that the population around Seoul has gradually increased from around 30% to 49.5% in 40 years.

In 1975, the total population in the Sudogwon area was just over 30%, then rose to around 45% in 1995. The figures continued to increase until it reached 49.5% in the 2015 census. It is therefore clear that people have constantly migrated to Seoul since 1975, and if these patterns continue, population of the *Sudogwon* is likely to exceed the *non-Sudogwon* population in the coming years.

These changes can be more explicitly demonstrated by their spatial patterns. In Figure 1.7, the population in 1985 was extremely dense in Seoul and in some districts of Incheon, a harbour city besides the Yellow Sea. The population density in Seoul was 15,921 persons/km^2 in 1985. In 1995, Seoul city decided to add three new districts to make 25 in total. At the same time, as the city's population density had already peaked at 18,000 person/km^2, the government built new towns in the southern and western outskirts of Seoul, in the expectation of a population decrease and in order to

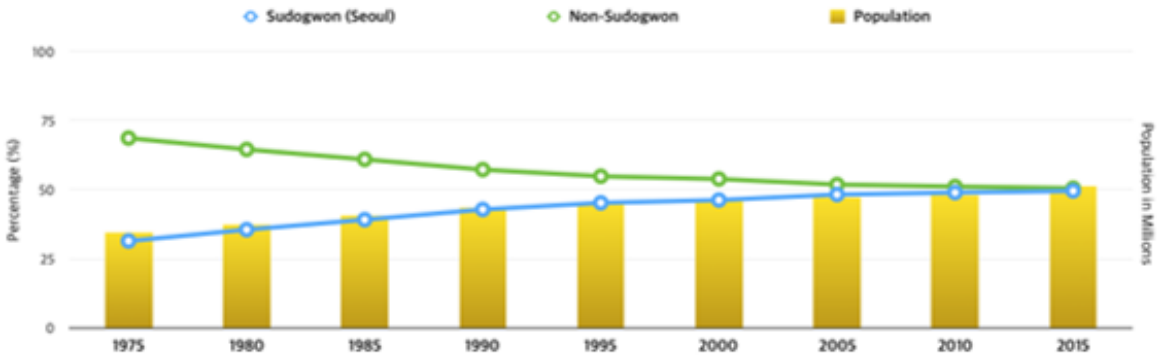


Fig. 1.6 Temporal patterns of population change between the *Sudogwon* (Seoul and the surrounding counties) and *Non-Sudogwon area*

reduce contaminants across the city. Despite the city's societal, economic, and cultural growth, it has managed to maintain a population density of around 17,000 persons/km^2 since 2000, by building new bed towns near Seoul, particularly in the south and west.

More migrants, however, from distant cities are moving to Seoul. This influx of population made several local governments and stakeholders concerned about investing new towns further outside. One idea was to construct new towns in the north and northeast of Seoul to help disperse the population, but unfortunately this action has achieved little success (Kwon et al., 2015). Understanding the dynamics of population trends is important for this project because the transformation rate from the statistics will be applied to evaluate the performance of a long-term simulation.

1.6.5 Air Quality Standards

The nation's atmospheric quality standards were created by the Ministry of Environment in 1978 to monitor sulphur dioxide. Since then, the standards had been revised six times by 2012, with pollutants expanded to eight categories in total. Notably, the importance of dust has been emphasized in recent decades, as coarse categories of total suspended particles (TSP) were once the only pollutant to be measured in 1991 but was removed in 2000 and replaced by PM_{10} and $PM_{2.5}$ in 1993 and 2012 respectively.

The Korean government has managed to reduce the average level of pollution concentration whilst attempting several environmental renovation plans. For example, temporal restrictions on traffic had been implemented in metropolitan cities during

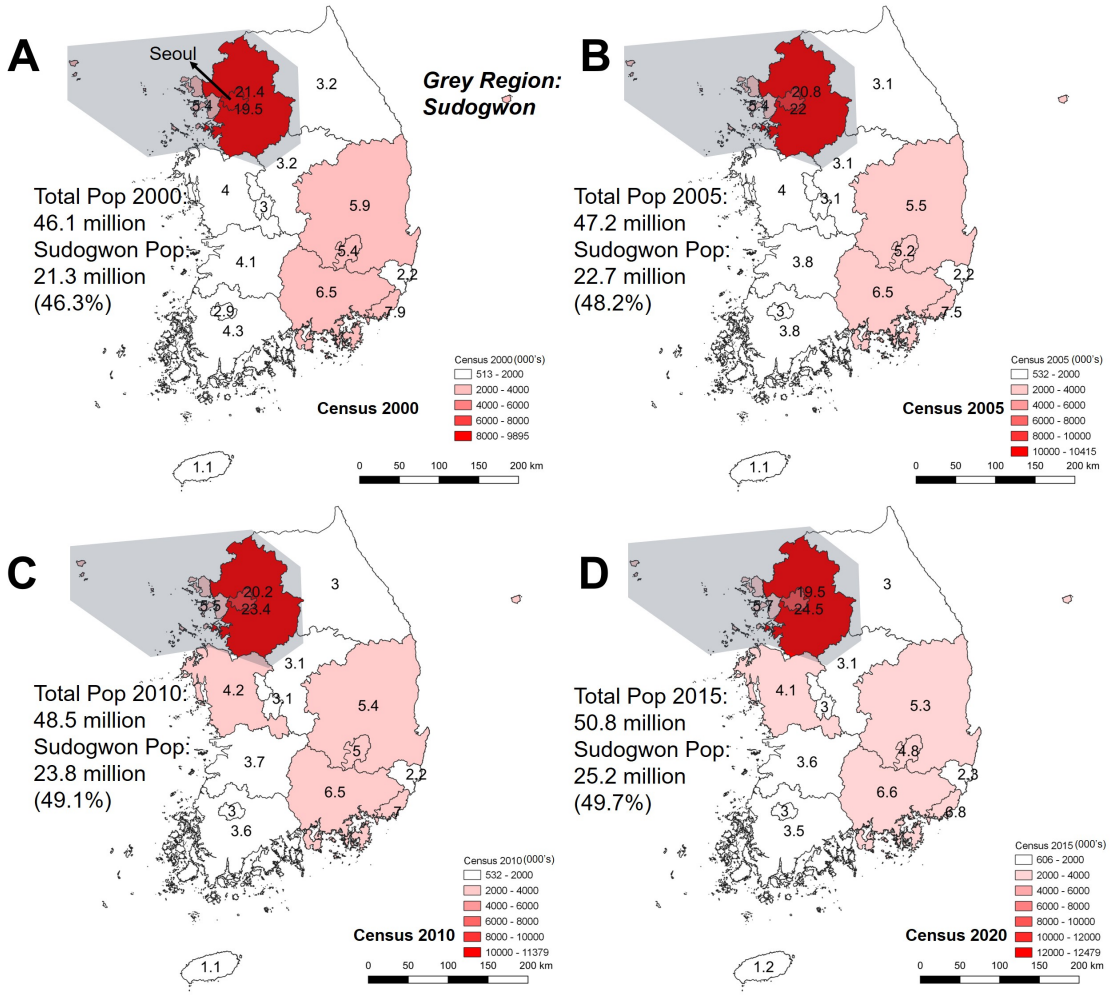


Fig. 1.7 Spatial and temporal change in population of South Korean provinces between 2000-2020. The shaded area (grey) in the northwest represents the *Sudogwon* area. This choropleth map illustrates the distribution of the actual population for each census and the numeric labels indicate the percentage of the total population. In line with Figure 1.6, more migrants from the *non-Sudogwon* cities have moved into Seoul and the neighbouring counties which led to a rapid population decline in other parts of South Korea. While the population in the Sudogwon area have increased, the population rate of Seoul have declined as citizens move out to the adjacent cities.

the 2002 FIFA World Cup (Lee and Kim, 2007) as per the low emission legislation of Seoul city. Nonetheless, these actions were less efficient to control local level pollution.

Many reasons have been raised about the causes of pollution rise. Initially, the pollution standards were more than twice as high as the WHO recommendations (see Table 1.3). There are arguments that the country's air quality standards are not as

tight as other European countries, which can miss the intermediate levels of pollution that can possibly harm vulnerable groups. Secondly, due to the fact that nearly half of the population reside on 10% of the land, pollution concentration and emissions are problematic. The Special Act on Seoul Metropolitan Air Quality Improvement has come into effect according to a presidential decree, which is leading the way to protect the health of residents in Seoul, Incheon and the Gyeonggi area. There are also claims that in recent years, a large amount of pollutants has been transported to Seoul due to emissions from the Chinese industrial complex on the east coast. The concerning fact is that the Chinese government has set a relocation plan for an industrial complex from Beijing to counties in the Shandong peninsula, which is only 250km away from Seoul.

This section provided concise but explicit reasons of why air quality is poor in Seoul with a description of standards in different countries. From a modelling perspective, it gives an idea to create possible scenarios that can help prevent deleterious effects on health, e.g. controlling pollution standards, or imposing congestion charges in the city centre. In the upcoming chapters, a mean of PM_{10} at 100µg/m^3 over the previous 24 hours will be used as a threshold parameter.

Table 1.3 Particulate Standard Comparison between Korea and Other Countries

Pollutant	Type	Korea	UK	US	WHO
PM_{10}	Annual Mean	50µg/m^3	40µg/m^3	-	20µg/m^3
	24 Mean	100µg/m^3	50µg/m^3	150µg/m^3	50µg/m^3
$PM_{2.5}$	Annual Mean	25µg/m^3	25µg/m^3	15µg/m^3	10µg/m^3
	24 Mean	50µg/m^3	-	35µg/m^3	25µg/m^3
NO_2	Annual Mean	30ppm	40µg/m^3	53ppm	40µg/m^3
	24 Mean	10ppm	-	-	-

Chapter 2

Literature Review

Chapter Outline

This literature review starts with a general introduction of pollutants in the atmosphere and how they are transported in urban space (Section 2.1). Section 2.2 documents the details of pollutants and their potential health effects, and how exposure patterns can vary the outcome. In addition, Section 2.2.3 introduces the main pollutants used in this research, which are PM_{10} and NO_2 that have the most sufficient evidence of causal association with adverse health outcomes. Section 2.3 constructs a conceptual framework that links the five key concepts of pollution, exposure, health risk, susceptibility and adaptation. The focus here is to not to critique the published studies but to better understand the relationship between the concepts. Once the concepts are established, Section 2.4 critiques how recent studies have quantified exposure both spatially and temporally – more specifically, how different toolkits are used in population and individual exposure research. Insights and gaps from the review suggest that agent-based modelling (ABM) is a suitable way to better understand the dynamics of exposure to air pollution for heterogeneous individuals. Finally, Section 2.5 outlines the methodological principles of ABM, its pros and cons, and a short review of how ABM and air pollution are applied empirically. Note that this chapter covers the core conceptual framework of the study, and more sophisticated reviews that target the aim of each study are documented within the following chapters.

2.1 Overview of Air Pollution

According to Britannica (2004), air pollution is defined as "a mixture of gaseous, fine solids or liquefied aerosols that exceed the environmental boundary of dilution, dissipation, or absorption". These substances are chemically classified in a dust form as particles, or a gas form as sulphur dioxides, nitrogen monoxide and dioxide, carbon monoxide, lead, methane, VOCs, and ozone. Most substances are generated from artificial emissions e.g. industrial sites, traffic, and domestic heating. Hence, people walking through these areas may be exposed to an unexpected level of pollution, and vulnerable groups are more likely to risk undesirable health problems such as pulmonary disorder, cardiovascular dysfunction, and eye irritation (Brook et al., 2004; Chen et al., 2007; IARC, 2013).

Most pollutants are generated in urban areas, and the level of pollution will increase strongly as urbanisation continues (see Figure 2.1). Megacities, including Seoul, are struggling with air pollution due to rapid growth of population and vehicles (Chen and Kan, 2008), and since South Korea has 25 million people (49%) living in and around Seoul, cars and building chimneys have contributed harmful pollutants to the atmosphere, which can further chemically react with other pollutants and deposit on surfaces (see Chapter 1 Migration Patterns for more information).

Air pollution can travel from a few metres to thousands of kilometres depending on the emission levels and weather conditions. A recent example was the Australian wildfires, which lasted for months between September 2019 and early 2020, where the fires released air pollutants like carbon monoxide, sulphur dioxide, nitrogen dioxide, and particulates, which contaminated the urban areas of Australia, some parts of New Zealand, and even South America. Malaysia also experienced a massive wildfire in 2015, where smoke from the forest was transported to Singapore and adjacent countries during the dry season. The wildfires whether natural or human-caused, make it evident that particles and other pollutants can be transported long distances and can potentially risk people's health in remote locations (Black et al., 2017). Therefore, locally experienced air pollution may be a mixture of long-range transported and locally generated material.

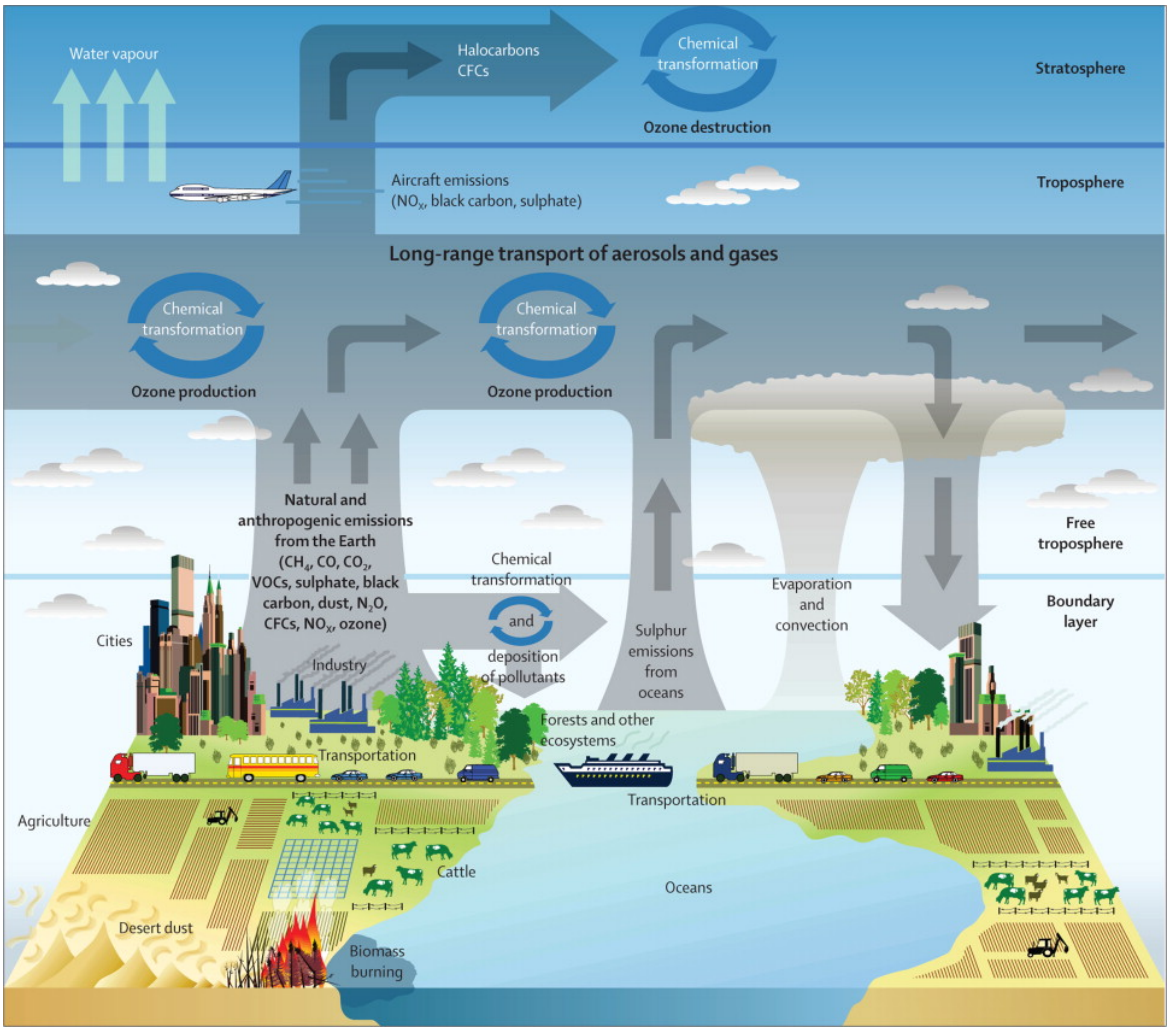


Fig. 2.1 Emission-transport-deposition circulation in the atmospheric environment (Guarnieri and Balmes, 2014)

2.2 Types of Air Pollution and Their Health Effects

2.2.1 Types of Pollutants

Atmospheric pollutants are categorised as gas or particulates. Korea's Ministry of Environment and the United States Environmental Protection Agency (US EPA) has established legal standards for three ambient gaseous pollutants, NO_2, SO_2, CO and for two ambient particulate pollutants, PM_{10} and $PM_{2.5}$. VOCs do not have a legal

standard because they mostly occur indoors but legislation was introduced to help understand there are other pollutants which can harm one's health.

Gaseous Pollutants

- NO_x: Nitrogen oxide is a highly reactive substance that is generated from fuel combustions of vehicles and power plants, and it becomes oxidative when it reacts with oxygen, which is known as NO_2. NO_x is currently the most harmful gaseous component reported in most urban areas globally.
- SO_x: Sulphur oxide is a substance that is mainly generated from fossil fuels in industrial sites, and the yellowish colour and the toxic odour are directly noticeable. Sulphur is also highly reactive to oxygen (SO_2) and it can be harmful to breathe. However, levels have dropped in Seoul since industrial sites have migrated to other areas or started to use renewable sources.
- VOCs (volatile organic compounds): VOCs are compounds of chemicals which include methane, carbon, ammonium carbonate, and so on, which have a high vapour pressure at indoor room temperature level (USEPA, 2020). VOCs are invisible but have a strong odour (e.g. paint). This, in the short term, can be negligible but can aggravate health in the long-term.
- CO (carbon monoxide): Carbon monoxide is a colourless and odourless gas type that is generated from the outside (e.g. vehicles, machinery, and burnt fossil fuel) as well as the inside (e.g. gas heaters, chimneys). While most developed countries have successfully regulated CO outdoors, developing countries and informal cities therein have struggled to meet standards.

Particulates

- PM_{10} (coarse particles): Particles are formed of numerous chemical substances, and usually the coarse particles primarily come from road dust, automobile engines (road and non-road), and secondarily from natural events such as wood burning, and dust transported from deserts. Coarse particles are easily recognised as a murky haze in an urban atmosphere as a result of stable airflow and a mixture with other chemical components.

- $PM_{2.5}$ or less (fine particles): The sources of finer particles are in indoor dust and cooking as well as unnoticeable components outdoors. In 2013, the IARC (International Agency for Research on Cancer) and WHO (World Health Organisation) declared outdoor particulates a Group 1 carcinogen (the worst substance group that causes cancer) because these particles penetrate one's lungs, bloodstream, and even brain which in turn deteriorates human organs (Hamra et al., 2014; IARC, 2013).

2.2.2 Health Effects of Pollutants

Prior to the section on exposure patterns, this section explores the potential health effects of exposure to pollutants mentioned above. Note that while some pollutants have evidence of short- and long-term causal relationships with health outcomes, others may only have long-term relationships (UK Department of Health, 2015; USEPA, 2020).

Gaseous Pollutants

Breathing gaseous pollutants can irritate the airways in the respiratory system. Studies have identified positive associations between short-term exposure to gaseous pollutants and hospital admissions for respiratory and cardiovascular symptoms (UK Department of Health, 2015). In particular, health effects from short-term exposure to NO_2 will result in nose and eye irritation (Idarraga et al., 2020) , coughing and wheezing, and for children, it can sometimes develop into acute bronchitis (see Figure 2.1) and pneumonia (Pershagen et al., 1995). However, how to define 'short-term exposure' might be controversial, because the effects were discovered either from 30 days (He et al., 2020) or 24 months (Chiusolo et al., 2011) of study period, which did not include exposure under 4 hours (European Commission, 2014). CO levels are rarely high outdoors but can surge indoors due to indoor garages, smoking, barbequing, or other burning activities. Short-term exposure to a high level of indoor CO will lead to dizziness and breathing problems. Short-term exposure to VOCs will cause nose, eye, and throat irritation (USEPA, 2020).

[1]Case courtesy of Dr Ian Bickle, Radiopaedia.org, rID: 50375

[2]https://www.openaccessgovernment.org/air-pollution-inflammatory-lung-diseases/67352/. COPD is a chronic/acute inflammatory lung disease that causes airway obstruction from the lungs. One of the common symptoms that appear on the X-ray is *emphysema*, which means the lungs get larger than normal. This inflation is attributed to many reasons but is likely to appear to children who are

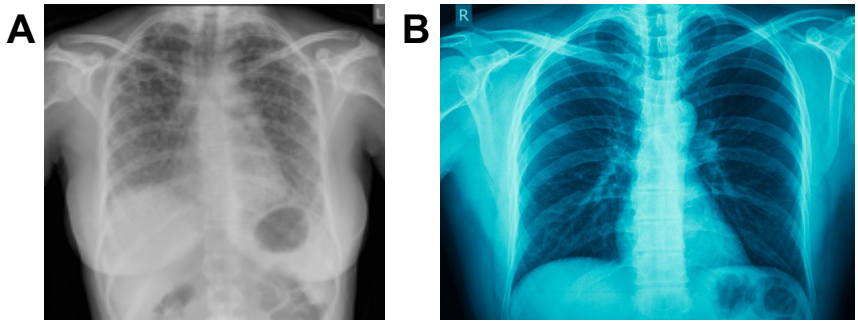


Fig. 2.2 Nitrogen exposure can result in (A) pulmonary fibrosis as a long-term effect[1], or (B) chronic obstructive pulmonary disease(COPD) as a short-term effect common in children[2]

Health outcomes from long-term exposure to NO_2 are pulmonary fibrosis[3] (see Figure 2.2), respiratory failure for children, and also mental illness. CO and SO_2 will also reduce the amount of oxygen that can be transported into the bloodstream, which will inevitably cause heart and brain malfunction (USEPA, 2020). In particular, long-term exposure to CO will lead to dizziness, unconsciousness, and death. A Chinese case study found that SO_2 had a strong association with asthma prevalence for children who had atopy (Dong et al., 2011); however, a more threatening aspect is that it contributes to form the particulates. Symptoms of long-term exposure to VOCs will cause nose and throat discomfort, continuous nausea, and sometimes dyspnoea (shortness of breath).

Particulates – PM_{10}, $PM_{2.5}$, and Ultrafine Particles

When a person inhales dust, it penetrates the upper respiratory tract and deposits itself in deep parts of the body depending on its size distribution and composition (see Figure 2.3). The reaction of the human body also varies with the dust's background source, whether it was anthropogenic or natural emissions, e.g. fungal spores or pollens (USEPA, 2020).

It is known that the health effects of short-term exposure to particulates are linked to respiratory and cardiovascular morbidity, such as acute asthma, respiratory problems,

exposed to (indoor/outdoor) air pollution. As pollution penetrates and damages the lung tissue, the airways will get obstructed and without treatment will likely lead to *emphysema*.

[3]Pulmonary fibrosis is a lung disease that happens when the lung tissue becomes damaged. The word fibrosis is named because the lung tissue thickens and becomes harder. In Figure 2.2A, multiple scars on the lungs indicate tissue damage in other words show how severe the symptoms are. As it worsens, the tissues will become harder leading to breathing disorder.

and eye irritation, but the consequence of long-term exposure to PM_{10} can lead to COPD[4], asthma, stroke, and cancer (Bhome and Brashier, 2014; Guarnieri and Balmes, 2014; Maantay, 2007). Recent studies have discovered that finer particulates less than $1\mu g/m^3$ are associated with cardiovascular symptoms as the small particles can enter the bloodstream (Cambridgeshire City Council, 2016; Guarnieri and Balmes, 2014; Loomis et al., 2013).

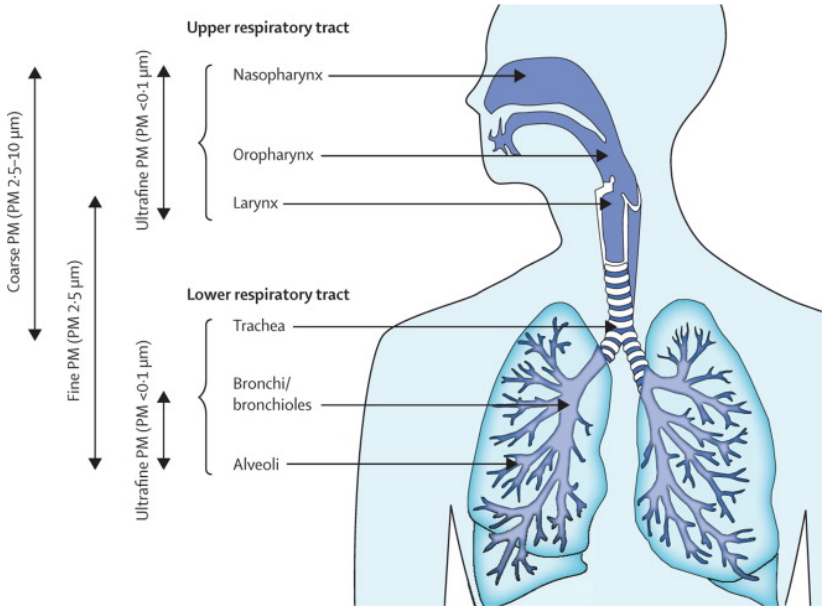


Fig. 2.3 Possible respiratory disease after particle inhalation (Figure imported from Guarnieri and Balmes (2014))

2.2.3 Pollutants Focused on in this Research

Amongst the many pollutants, this study focuses on PM_{10} and NO_2 for the empirical chapters. PM_{10} was selected because it has been recognised after the finding that the increasing level of PM_{10} is highly correlated to lung cancer – outdoor particulates are carcinogenic to humans, and thus is categorised with Group 1 carcinogens (IARC, 2013). NO_2 is also a detrimental pollutant affected by traffic levels and also generates ground-level ozone levels which can create respiratory problems in humans. The causality between pollution and the outcome has been proven in journals and governmental documents and are still the most frequently analysed components in the literature.

[4]A progressive disease that makes it hard to breathe (National Institutes of Health)

PM_{10} is the main pollutant focused on in Chapter 3, 4, and 5. NO_2 is also the focus of in Chapter 3.

2.3 Contextualising Air Pollution Exposure and Health Risk

This section constructs a conceptual framework to better understand the relationship between air pollution and health risk. Parry (2007) defined vulnerability as "the susceptibility of individuals to the impacts of hazards as well as their resilience to recover from them", which was originally one of the key concepts of climate change. As mentioned in Chapter 1, vulnerability is composed of exposure, sensitivity, and adaptive capacity. A similar concept was discovered in spatial epidemiology by Jerrett et al. (2009), where the authors provided a diagram of 'health risk' from air pollution exposure as a Venn diagram. For this study, the Venn diagram was converted to a process diagram to illustrate how one's health risk results from the exposure pattern to air pollution (exposure), the variability in innate status or societal status (sensitivity/susceptibility), and personal resilience or the regional environment which helps mitigate the risk from air pollution (adaptation) (see Figure 2.4). While Parry (2007) and Jerrett et al. (2009) suggested a conceptual and deterministic approach respectively to account for the relationship between the main effects, there are, in reality, uncertain factors that can result in health risk that include the time spent outside and mobility patterns, which interact in a complex way.

In this regard, this thesis takes a step forward by asking the questions, "how can different exposure patterns translate to risk?", "how sensitive is the outcome to behaviour and demographic status?", "what is the personal/regional reaction to adapt or cope with air pollution?". The first question is answered in the following section by examining various exposure patterns both conceptually and empirically. The next section explores further the susceptibility of individuals relating to their societal status and demographic groups and within these groups, and how behavioural patterns can also be important to address susceptibility. The final section attempts to understand the conflicting concepts of adaptation and susceptibility which have been used in exposure studies.

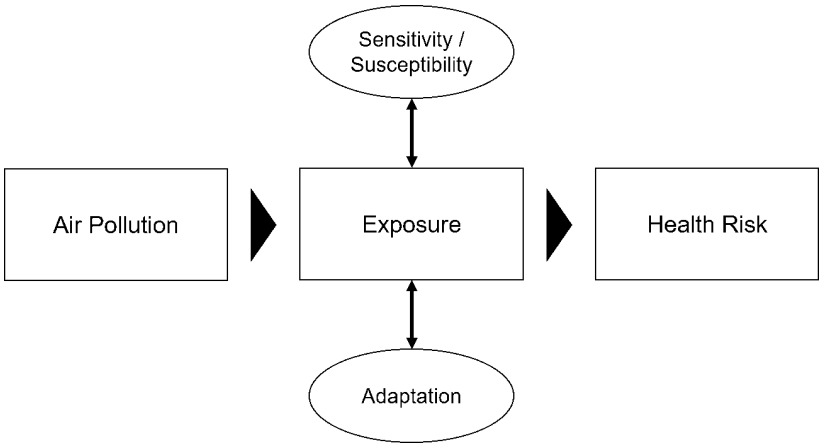


Fig. 2.4 Conceptual model for the research project (developed and converted from a diagram drawn by (Jerrett et al., 2009))

2.3.1 The Relationship between Exposure and Health Risk

To answer the first question "how can difference exposure patterns translate to risk?", this section describes how various exposure patterns can lead to different adverse health outcomes. Examples of exposure patterns in Figure 2.5 include: (a) continuous, (b) intermittent, (c) cyclic, (d) random, and (e) concentrated (Deshpande, 2002). These patterns can appear in a short-term cycle from hours and days or a long-term cycle to months and years. This means that one can have severe health issues having been exposed to a constant level of intermediate pollution or resulting from myriad experiences of extreme pollution episodes (Laumbach et al., 2015).

Conceptually, the association with health risk is determined by how frequently people are exposed to ambient air pollution especially when the levels exceed the national standards, either short-term or long-term (Deshpande, 2002). There is a higher chance for people to be exposed to a detrimental level of pollution during weekdays as people commute to their offices at the same time. Weekday commuters are likely to be exposed to intermittent or sudden peaks of pollution caused by heavy traffic congestion, road dust, and indoor resuspended particles. In addition, exposure can depend on the time of the day in which the individual was exposed. In the literature, highest exposure has been reported in the morning rush hours (06h-09h) where most of the traffic congestion is seen in the business districts (Beevers et al., 2013; Cariolet et al., 2018; Vreeland et al., 2017; Zhang and Batterman, 2013). European cities have introduced vehicle restrictions to curb air pollution from exceeding the WHO standards;

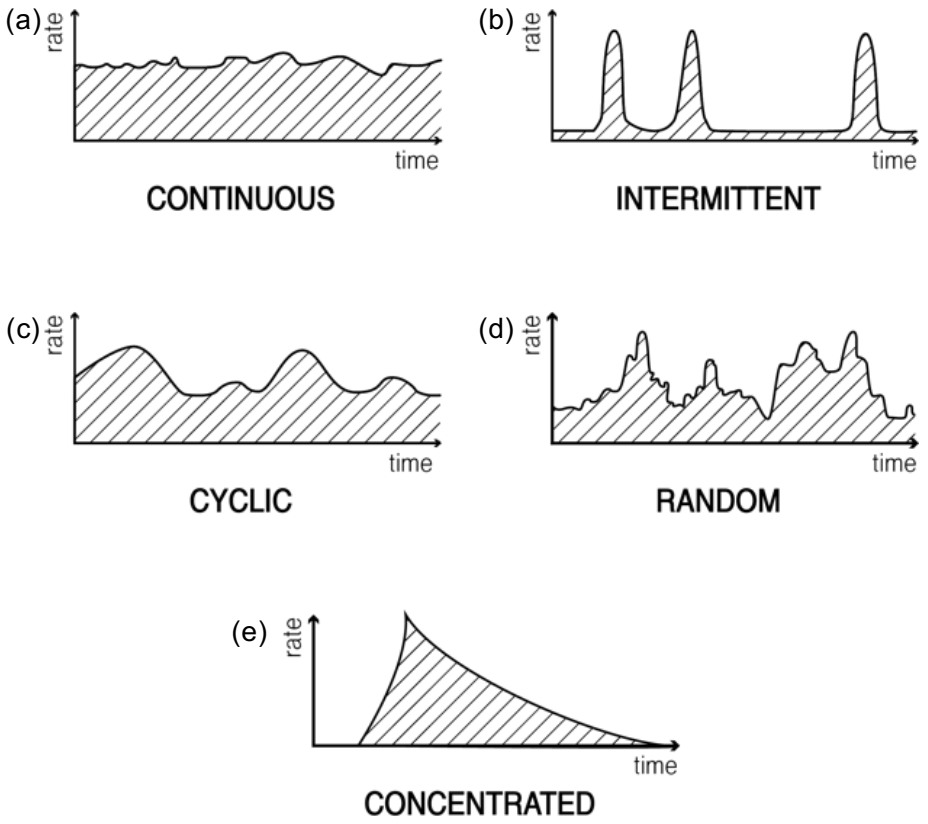


Fig. 2.5 Patterns of Pollution Exposure (redrawn from Deshpande (2002))

nonetheless, those who drive in rush hours have higher health risks due to the harmful pollutants that can ingress inside vehicles (Kumar et al., 2020).

Furthermore, exposure also depends on which place an individual is located. In urban areas, city centres have the highest population density and traffic volume during working hours. Cariolet et al. (2018) used a 500m×500m resolution map of greater Paris and found that inner Paris had the least capacity to reduce NO_2 concentrations due to high building densities and higher traffic. A study in London found that the city centre was not the hotspot of black carbon (BC) and NO_2, rather the surrounding roads that lead to the city centre had greater pollution (Rivas et al., 2017). Kobza and Geremek (2017) also found that motorways and urban expressways had experienced a higher NO_2. Subways were also known as a hazardous area for increased health risks. In London, Smith et al. (2020) monitored $PM_{2.5}$ levels between Waterloo, and alternating destinations of Oxford Street (high traffic location) and Hyde Park (low traffic location) via the Jubilee Line for five months. The study revealed that the

underground $PM_{2.5}$ was around 89µg/m^3 on average (median 28µg/m^3), which was greater than at ambient background locations (mean 19µg/m^3, median 14µg/m^3) and the roadside in central London (mean 22µg/m^3, median 14µg/m^3); however, there was significant variability between the shallow and deeper stations with an average of 4µg/m^3 and 361µg/m^3 respectively. A similar study has been conducted for Seoul, where PM_{10} and $PM_{2.5}$ were measured from subway platforms and carriages at 22 stations on Line number 1-4. Between November 2004 and February 2005, the study discovered that the average level of PM_{10} and $PM_{2.5}$ exceeded 300µg/m^3 and 120µg/m^3 at the platform and inside carriages, while the driver's seats also had 271µg/m^3 and 127µg/m^3 respectively. High exposure was also reported near schools during drop-offs and pick-ups (Kim et al., 2004; Kumar et al., 2020; Min et al., 2020).

Finally, exposure is highly dependent on meteorological and infrastructural components. Exposure patterns are known to be substantially lower when the wind speed is higher because the wind blows away the deposited pollutants, but when the wind speed is low, the airflow tends to trap the flow of the pollution in tall building areas, which in turn increases the pollution levels (Cariolet et al., 2018). In these areas, vehicle emissions, either exhaust or non-exhaust, can disperse to roads and nearby paths which can harm both the in-vehicle passengers and pedestrians.

In practice, many studies have found a causal relationship between exposure and adverse health outcomes, mostly through long-term investigation of patients whose diseases were attributed to air pollution (Halonen et al., 2016; Hoek et al., 2013; Kaur et al., 2005). These studies found the relationship by monitoring the chart history of patients from the start of the symptoms to drug type and dosage, days of hospitalization, and the discharge from hospital. Even if a small sample was collected for the analysis, the outcome was still meaningful. However, the government reports suggest that more evidence needs to be collected to associate the causality between short-term exposure and adverse health outcomes because short-term morbidity or mortality can be affected by other hidden confounding factors such as alcohol ingestion and smoking (European Commission, 2014; Guarnieri and Balmes, 2014; UK Department of Health, 2015).

The WHO (2003) and Atkinson et al. (2016) also supported the idea that long-term assessment provides more accurate results in terms of health outcomes. Atkinson et al. (2016) compared both the short and long-term (722 days) exposure to NO_2, $PM_{2.5}$, and black carbon (BC) in London, and discovered that there was a weak, but positive association between long-term exposure and respiratory mortality (but not

with cardiovascular mortality); however, the study found no association with short-term exposure to NO_x or NO_2. Indeed, there are uncertainties and other auxiliary factors that may lead to death and this cannot be causally related in a short period (UK Department of Health, 2015). Also, using station data to model pollution can differentiate the locations of where people were exposed.

Although a few studies still support the health effects of short-term exposure arguing that evidence of a response to acute pollution is present, particularly to cohorts of asthmatic children and adults who are exposed to outdoor $PM_{2.5}$ and $PM_{2.5\text{-}10}$ (Cambridgeshire City Council, 2016; Guarnieri and Balmes, 2014), it is difficult to conclude whether health outcomes are more detrimental when one is exposed to an intermediate pollution level over a long period, or to several events of instantaneous peaks. People who are already made vulnerable by chronic exposure over a long period may then experience a sudden effect owing to an acute incident.

2.3.2 Associating Exposure and Sensitivity with Health Risk

Sensitivity[1] relates to the characteristics of individuals whereby the negative impacts of air pollution can appear unequally. The determinants of sensitivity can be categorised as biological sensitivity and environmental sensitivity. Biological sensitivity can be age, sex, genetics which are innate components, while acquired components include diet, smoking, and disease history (Holgate, 2017). Environmental sensitivity includes housing address, type of house, distance to roads/industrial sites, commute patterns, and occupational surroundings (Holgate, 2017). Amongst these determinants, this study answers the question "how sensitive is the outcome to behaviour and demographic status?" by documenting essential parameters that had a causal association between pollution and health risk.

Initially, groups vulnerable to air pollution are more likely to be children and the elderly (Cambridgeshire City Council, 2016; O'Neill et al., 2003). Children are the most susceptible group out of three age classes: 0-14, 15-64, 65+. Sacks et al. (2011) and Pearce et al. (2006) suggested that higher morbidity/mortality for children is more likely due to the children's immature physical system and many hours of outdoor activity. For example, in the UK, a 9-year-old girl living near the M25 motorway in south London suffered and died of an acute asthma attack: this was a landmark

[1]Here, the term sensitivity and susceptibility can be used interchangeably

case of nitrogen oxide (NO_x) directly associated with fatality (Hancock, 2020). Wang et al. (2008) discovered that population-weighted exposure (PWE) of elderly groups was highest among all demographic groups because these groups have retired from labour and have more time spent indoors, and perhaps exposure to solid fuel heating in developing countries.

Additionally, studies found that people of lower socioeconomic status (SES) tend to be vulnerable to air pollution (Batterman et al., 2015; Dons et al., 2014; O'Neill et al., 2003; Sacks et al., 2011). Environmental justice researchers have shown the existence of a correlation where people who live in disadvantaged neighbourhoods have a higher risk of a health disorder due to noise, pollution concentration, and lack of green areas (Mitchell and Dorling, 2003; Moreno-Jiménez et al., 2016; Pearce et al., 2006); this is also strongly linked to race (Loo et al., 2017; Mitchell and Dorling, 2003; Park and Lek, 2016; Pearce et al., 2006; Sacks et al., 2011), income (Guarnieri and Balmes, 2014; Hoek et al., 2008; Sacks et al., 2011), and education (Bravo et al., 2012; Jerrett et al., 2001; Kan et al., 2008). Dons et al. (2014) pointed out that low SES (Socioeconomic Status) people inhaled an average of 465ng/day more than that of high SES because these people had less of a commute distance and might have stayed in rooms where the air filtration rate is low (Ott et al., 2006). Professor Jamie Pearce has worked on the geographic features of social inequality for a long period, and through his outputs of environmental justice in the case of New Zealand, the lower socio-economic deprivation quintile was substantially associated with high PM_{10} exposure (Pearce et al., 2006, 2011).

Infrastructural factors have also continuously been associated with health effects. Guarnieri and Balmes (2014) highlighted that the concentration of many constituent pollutants rapidly diminishes with distances within 500m of roads in North America, and 300m in European cities. Van Roosbroeck et al. (2007) measured personal exposure to traffic-related air pollution in Utrecht, the Netherlands. For 54 young participants, the personal exposure to NO_x was 37% lower in near-road school children compared to those who went to school further from main roads. Kim et al. (2004) conducted 10 school-based surveys on students who walk to school to compare NO_x exposure levels based on their residential location in San Francisco, USA, using a logistic model. Children who lived within 300m to major roads and at a downwind location had higher exposure levels (OR: 1.05 ± 0.1) compared to those who live far from and upwind of major roads.

In short, sensitivity is a very important parameter because the health conditions not only depend on the person's innate status (e.g. sex, age, being inherently allergic to some chemical products) but also on the societal status (e.g. occupation, residential area, type of housing, income). If this study assessed exposure at a regional scale, then the "regional averages" of education, age, or income would have been used to measure health risk. However, as this study intends to quantify the exposure levels at an individual level, the levels would vary substantially within the groups. Moreover, whether one has a history of disease or which behavioural patterns they have also translates to a different health outcome. Obtaining a finer scale can be difficult but it is important to collect as detailed information as possible because creating unique profiles for every single individual will allow a firm foundation to create more realistic outcomes.

2.3.3 Relating Adaptation to Exposure

Adaptation (or resilience) is how an individual or society can adapt to the impact of air pollution. Adaptation has been applied in both individual and regional level research.

An individual level of adaptation from air pollution exposure can be seen through physical activities. Recently, the Centre for Diet and Activity Research (CEDAR), part of the Epidemiology Unit at the University of Cambridge, has found that the health benefits of cycling and walking can outweigh the aggravated effects of air pollution (Tainio et al., 2016). In this study, the authors compared the exposure levels between walking and cycling participants as a representative group of physical activity and driving participants as a group of inactivity. Having collected the data from many cities around the world, the study revealed that cycling and walking for 45-60 mins daily can outweigh health risks from air pollution compared to driving (inactivity). A similar study used Cox proportional hazards regression to understand how physical activity can help lower the risk of long-term exposure to $PM_{2.5}$ for the elderly (Sun et al., 2020). The study found that cardiovascular death risks were negatively associated with physical activity, particularly for those who were doing indoor spinning and Tai Chi (Chinese exercise) classes, although a $10\mu g/m^3$ increase in $PM_{2.5}$ was associated with a hazard ratio[2] of 1.19 (95% CI: 1.05, 1.35) for cardiovascular mortality. Both

[2]The hazard ratio is a comparison of the probability of events in a case group, compared to the probability of events in a control group. If the ratio is above 1, then it means that 1.19 times as many patients in the case group are experiencing an event compared to the control group.

of these studies have provided insights that exercise can help lower health risks even in highly polluted areas, but careful consideration is needed for people who travel or exercise less than 30 minutes per day or take other modes of transport. One thing to note is that these studies have only taken a few participant samples and controlled the activity patterns, which does not monitor the behaviours for a continuous period. This means that people might possibly have improved or worsened their health through other activities.

From a regional aspect, the level of adaptation has been assessed based on land use types. Panko et al. (2013) investigated the relationship between potential human exposure to non-exhaust emissions, mainly from tyre and brake wear. The authors collected air samples from roadside, residential, commercial, school, urban, rural and recreational settings from selected cities in the US, France, and Japan. They discovered that the contribution from tyre and brake wear was less than 1µg/m^3 on average, and areas that had more recreational spaces (e.g. parks) and less major road sources had fewer particles. Cariolet et al. (2018) assessed the resilience of urban areas to traffic-related air pollution in Greater Paris. Using raster grids to calculate the capacity to decrease exposure to NO_2, the authors revealed that 6% of grid squares in Paris have continuous exposure to NO_2 and most of the inner Paris region currently has a low capacity to decrease exposure due to having such closer distances to roads that have more than 10,000 daily users. Also, the building density in inner Paris disturbs the wind which blows away the pollution levels, thus suggesting that more building-level ventilation is needed. Xiong et al. (2019) assessed how social resilience will potentially mitigate the health risk from air pollution in Sichuan, China. Resilience parameters included hospital density, GDP, roads, and land use. The study found that the regions with grazelands accounted for higher social resilience, although the hospital density was very low. These results can be useful to understand how administrative regions are capable of mitigating risks from air pollution, but on the other hand, it implies that a few weighted factors, such as land type, can over-determine the outcomes, which might not be true at the individual level.

From the review above, adaptation at an individual scale has provided a solution to mitigating health risk from air pollution by taking action with one's fitness: exercise is good even in high air pollution. Although it depends on the frequency and the distance between home and work, this is a useful indicator that can be used in simulation studies. Regional-scale adaption does not offer as many dynamics of human interactions and

resilience than individual-level outcomes. This is because health risk is evaluated by land type features or by gridded outcomes, which will allow a holistic perspective. Since the focus of this study is to estimate population exposure to air pollution from an individual-level simulation, it might not be feasible to consider physical activity by every individual to test the resilience, but it can apply ways to downscale the regional components to an individual level, for example assuming access to the hospital is higher/lower for individuals in a certain area.

2.3.4 Summary of the Section

This section contextualised the process of how spatial and temporal patterns of exposure to air pollution can lead to health risks. From the review, it is understood that air pollution itself appears to spatially vary by metres to kilometres, and temporally vary by minutes to years, yet more importantly, varies by how long one is exposed to (the ambient levels of) pollution, and how much time is spent at that place. The chapter also notes that even though the exposure might be equivalent between individuals it can be very sensitive to the innate health condition of an individual as well as which demographic/societal group the person belongs to. For adaptation, physical activity was used in individual adaptation while land-use types, wind speed, ventilation, and hospital access were selected for regional adaptation.

2.4 Exposure Assessment Methods: From Spatial Interpolation to GPS Tracking

To date, the mainstream topics in exposure science include the keywords ‘population exposure’ or ‘individual/personal exposure’. This section reviews studies with a focus on the methodologies that have been used to evaluate ‘population’ and ‘individual’ level exposure, how different spatial and temporal resolutions are, their pros and cons, and finally will discuss the gaps which need improvement.

2.4.1 Spatial Interpolation

Before the 2000s, it was very common to get uniformly interpolated air pollution (using nearby monitoring data) across a spacious area when fewer monitors existed. Obtaining nearby pollution data was effective in small area studies where the monitoring station was close enough to examine, for example, school children's exposure to particulates (Miller et al., 2007; Moshammer et al., 2006), but it became obsolete as more stations installed in cities gave more detailed information about pollution. Still, this 'nearby' method has been used as a reference method in spatial interpolation studies (Son et al., 2010; Wong et al., 2004; Xie et al., 2017).

Spatial interpolation methods have inferred unknown locations based on their arithmetical algorithms. Studies have applied Inverse Distance Weighting (IDW), Kriging (Ordinary, Universal, Block), and Generalised Linear Models from the monitor data (Min et al., 2020; Son et al., 2010; Wong et al., 2004; Xie et al., 2017). These methods used an annual or seasonal average of pollution to predict the generic level of pollution over a large surface area by weighting algorithms.

From a spatial aspect, the accuracy of air pollution in the predicted areas was determined by the density of the fixed-site monitors. This is because spatially dense monitors can provide more information to reduce the error estimation that can happen if the monitors are not present. However, if the distribution of the stations is biased to particular zones (i.e. violating the assumption of stationarity), then it may be difficult to get accurate predictions in data-sparse areas. The biased distribution of monitoring stations clustered in California and Northeastern regions in the US can be used as a good example of why data biases can be problematic for pollution exposure studies. From a temporal aspect, interpolation studies used annual mean pollution to predict population exposure. This is because the seasonal cycles of pollution, particularly for PM_{10} and $PM_{2.5}$, are very distinctive compared to other pollutants, but it can be pointed out that an annual average will smooth away any real variability which can be hazardous to health.

Min et al. (2020) have associated school children's exposure to annual NO_2, PM_{10}, and $PM_{2.5}$ with allergic atopy symptoms. With Universal Kriging, the outcomes have found a positive relationship between traffic and allergic symptoms in children as the schools were closely located to roads. However, as atopy more likely occurs during seasonal transitions where the range of daily temperature is high, the annual

averages will not fully articulate the relationship between the disease that happens to be monitored day by day. Also, the exclusion of built environments such as roads, buildings, and parks will cause a large discrepancy to model accurate numbers. The London Atmospheric Emission Inventory (LAEI) has measured NO_2 exposure in 2016 at a 1km by 1km spatial resolution and counted potential health harm by boroughs that exceeded 40µg/m^3 (Greater London Authority, 2017). Greenwich, for instance, had 22,900 people (8%) potentially exposed to high NO_2. However, in central London, the population-weighted exposure to $PM_{2.5}$ in 2008 was 28% lower when the building's indoor/outdoor ratio was considered compared to the method that simply used ambient $PM_{2.5}$.

Note that this section does not introduce the development of spatial interpolation models but will compare Kriging and Generalised Additive Models (GAMs) in Chapter 3.

2.4.2 Dispersion Models

Atmospheric dispersion modelling is a mathematical and computational method that simulates the concentration of air pollution by calculating chemical physical processes of plumes, and dispersion mechanisms (Holmes and Morawska, 2006). Dispersion modelling has its significance in pollution studies because pollution monitoring sites only collect point data, which cannot sufficiently estimate concentrations over wide areas, Whereas models can give results at multiple scales, from street canyons to a country scale, although some degree of validation is necessary (Barratt, 2013).

Dispersion models also help make a better estimation of pollution concentration, since the models take into account various weather events: plume generation and convection, building downwash, meteorology, and topographical conditions. Published studies from an interpolation approach have yet to consider much about built environments and climatic effects, despite sophisticated algorithms and a variety of tools.Compared to earlier studies, fewer geostatistical interpolation methods have been used recently because a consensus has not been made for a better way to make estimations in air quality modelling (Bravo et al., 2012), and due to the difficulty accounting for road network and buildings.

There are a number of dispersion models that researchers can choose based on the spatial scales. If the study examines data at a street canyon scale (<500m resolution),

then the Gaussian plume models are adopted e.g. CALINE[3], ADMS[4], or R-LINE[5]; however, if the study examines data at a mesoscale area (>5km), then an Eulerian model (three-dimensional gridded transport system) is considered e.g. CMAQ[6], TAPM[7]. A detailed explanation of ADMS, CMAQ, or TAPM is covered in the Appendix. Gaussian plume models take advantage of pollution concentration in near-field situations, which might be from a few hundred metres distance to a few kilometres (Williams et al., 2011), but lack the mixture of chemistry conditions that commonly happens in the local atmosphere, which is important in urban scale modelling. Thus, for a city-scale dispersion, there is a need for background level concentrations from regional models.

Using CMAQ, previous studies have also discussed the relationship between health impacts and ambient air pollution, particularly ozone (Bell, 2006; Knowlton et al., 2004), both ozone and $PM_{2.5}$ (Beevers et al., 2013; Smith et al., 2016). Reviewed research was conducted in the USA (mainly northeast), the UK, and Korea, but other studies have been carried out in almost 50 countries.

Knowlton et al. (2004) investigated the summer ozone concentrations and their relation to mortality, with CMAQ's dispersion model SMOKE[8] and weather forecast model MM5[9]. The authors concluded that the rise of ozone projected an increase in the mortality rate by around 4.5% for New York by 2050 if the USA continues to generate equal amounts of airborne emissions from power plants or continues to over-use air conditioners or dispose waste at landfill sites. Similar work was conducted by Fann et al. (2012), where CMAQ was used to estimate $PM_{2.5}$ and 8-hour ozone levels with a 12km grid cell, and examine the relationship between statistics of premature death and life expectancy in the United States. The result predicted 130,000 deaths to $PM_{2.5}$, and 4,700 to ozone and that over 1.1 million elderly people are more likely to have their life expectancy reduced. Also, Fann et al. (2012) gave an interesting rank of the mortality rate of cities that were affected by ozone and $PM_{2.5}$ in 2005. However, these results might have over-predicted risks in terms of interpreting pollution

[3]a line source air quality dispersion model that is based on the Gaussian diffusion equation and employs a mixing zone concept to characterize pollutant dispersion in the proximity of roadways.

[4]Atmospheric Dispersion Modelling System

[5]A Research LINE-source dispersion model for near-surface releases

[6]CMAQ is a three-dimensional air quality model to estimate concentration of multiple pollutants at different geographic scales (Beevers et al., 2013; Bravo et al., 2012; Hoek et al., 2008). CMAQ, pronounced 'see-mak', was conceived with the intention of building a 'one atmosphere' and a high-quality model which can help to observe air quality over a long time series (Williams et al., 2011)

[7]The Air Pollution Model

[8]Sparse Matrix Operator Kerner Emissions

[9]mesoscale meteorological model - 5

concentration associated with mortality figures. Hence, more consideration is needed before examining the link between mortality and air quality standards.

Some epidemiology studies used CMAQ without putting health measures forward (Bell, 2006; Bravo et al., 2012). These studies used CMAQ modelling to fill in area gaps for estimating exposure levels where monitoring stations do not exist. These studies were possible because CMAQ improved the spatial resolution down to 1km, which can verify the data with more than 2000 stations. Hourly-based time series also gave a detailed pollution pattern. Bell (2006) estimated exposure levels with multiple approaches, including area-weighted average pollution using CMAQ, nearest monitor, and spatial interpolation (i.e. IDW, Kriging). They concluded that previous methods gave a poor value at places where monitoring stations were sparse, and suggested the need to thoroughly examine modelling methods before measuring personal exposure. Bravo et al. (2012) also simulated pollution dispersion by aggregating the values from 1km grid cells and compared the modelled values with populations where monitoring stations are either close or not. Results showed that people who reside within 50km of the station have higher incomes, higher academic education, higher employment rates, and a lower poverty index.

In short, CMAQ has strongly contributed to pollution and health studies with respect to technical improvement. However, area-weighted average values in 1km grid cells are too coarse to detect the dispersion concentration on an intra-urban scale, and suggests where new measuring stations should be installed, either at a regular distance or population-derived measurements (Song and Kim, 2016).

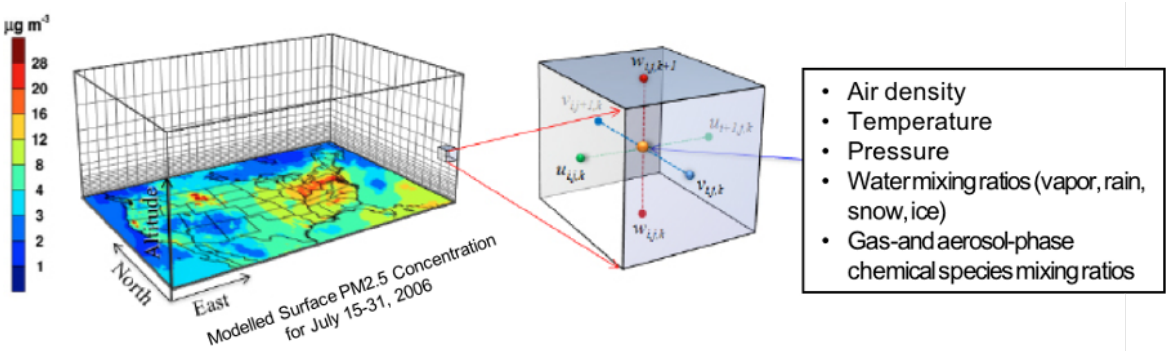


Fig. 2.6 CMAQ grid cell features (http://bioearth.wsu.edu/cmaq_model.html)

To overcome the limitation of coarse spatial resolution, a new hybrid model of city-scale dispersion was constructed by combining a street canyon model, ADMS.

King's College London's epidemiology team developed KCLurban, where they combined the US EPA's CMAQ model and ADMS-Road as per the need to estimate individual exposure rather than from fixed monitor stations (Beevers et al., 2013; Smith et al., 2016). Despite the existence of KCLurban as a UK standard model of NO_x and O_3, Beevers et al. (2013) suggested a new model that takes into account a hybrid of regional and road dispersion models, with a mesoscale weather forecasting model (WRF v3.1). With KCLurban, the result between predicted and observed values of NO_2 and PM_{10} at 16 monitoring sites showed a strong R^2 at 0.93, and 0.96 respectively.

A project by Smith et al. (2016) constructed a city level exposure model for the Greater London population to assess air pollution sources that affect outdoors, indoors, and in-vehicles using time-activity data from surveys of 8,000 households per annum between 2005-2010, and a trip route simulation from 45,079 people taking 98,770 trips. Using a confirmed CMAQ-urban model, the output was given as an annual average of NO_x, NO_2, and $PM_{2.5}$ concentrations in a 20m $\times$ 20m resolution map across the UK on a single day of 2011. The authors found that over 95% of people spent most of their time indoors, and that time affects the population exposure up to 92% for NO_2 and approximately 95% for $PM_{2.5}$ from outdoor effects. They gave a clear idea of exposure levels by using a sample number of residents' addresses to reduce biased records, but the activity model is only recorded for a day, which could still produce biases.

Cambridgeshire City Council (2016) also used ADMS to report adverse health impacts from air pollution in Cambridgeshire with the 2016 data (see Figure 2.7). The council estimated 257 deaths (5% of all deaths in Cambridgeshire in 2010) were attributable to air pollution in Cambridgeshire, particularly the elderly and children, who live near or pass through the city centre (Drummer Street) or trunk roads (A14) and are more likely to experience risks due to traffic congestion. Although the dispersion of PM_{10} is not shown in the figures, it was reported that an increasing trend was seen between 2009 and 2013 in A14 roads in Impington which may have been affected by traffic, and Cambridge city had the highest monthly mean concentration in March.

Although ADMS considered meteorological factors and vertical convolution effects to provide high-resolution and accurate results, there might be a possible downside. Firstly, it is too slow to model a city with an ADMS-Urban package, most likely due to the model's single-core design, or an unsorted algorithm when blending two different

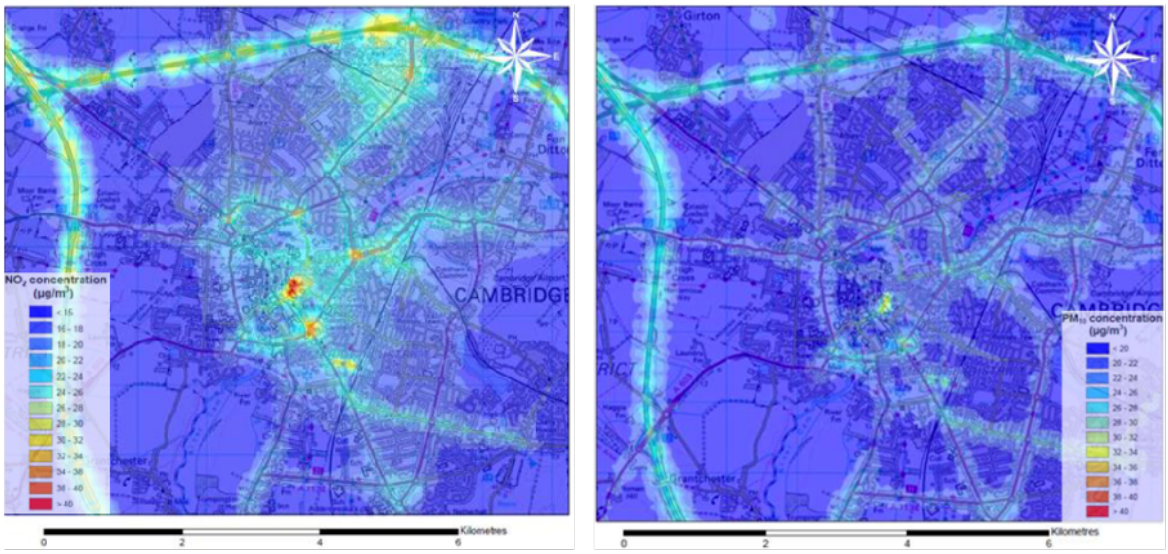


Fig. 2.7 Annual average values of NO_2 and PM_{10} 2016 in Cambridgeshire using ADMS-Road

packages[10]. Also, ADMS is for commercial use only, which means a certain amount of money needs to be paid (at least £3,000) for the annual license plus tutorial classes.

2.4.3 Towards Personal Exposure: GPS and Mobile Phone Data

Estimating population exposure with activity patterns has been introduced to improve the accuracy of location and exposure levels both in space and time. These activity patterns are demonstrated with surveys, OD (Origin-Destination) matrices, GPS, and smartphones.

Smith et al. (2016) used time-activity data based on the London Travel Demand Survey (LTDS) to model the movements of Londoners. KCLurban was rebranded as the London Hybrid Exposure Model (LHEM). Having taken surveys of 8,000 households that approximated 98,770 trips for a single day in 2011, the average exposure to $PM_{2.5}$ was 37% lower, and NO_2 was 63% lower in the survey results than merely calculating ambient concentrations based on residential addresses. The survey-based results seemed to be a plausible approach because the survey found out that 95% of the participants

[10]According to Williams et al. (2011), it was revealed that the structure of ADMS-Urban was a blend of ADMS-Road and ADMS 5.

spent most of the time indoors during the day, and the indoor/outdoor ratios varied by 15 building categories, from houses to schools and institution buildings.

Using 5-million mobile locations from residents of Belgium[11], Dewulf et al. (2016) estimated exposure to hourly NO_2 maps that were produced by the RIO-IFDM model[12]. With a full two days of pollution estimation, it resulted that compared to the static census data, the mean exposure to NO_2 per person from mobile phone data was 1.2µg/m^3 higher. Nyhan et al. (2016) also compared census data to an estimated population data generated by 3G Mobile Networks in New York City. Although $PM_{2.5}$ was very coarsely interpolated with IDW, over 10 districts in NYC was appeared to be higher in $PM_{2.5}$ exposure than the census result. However, Xu et al. (2019) criticised that mobile phone data only displays the location of a population in a given grid provided by the cell-tower data, which will allow some representation of movement but not for unique individuals. Larkin and Hystad (2017) also pointed out that the representation of the data can strongly depend on the mobile company and whether the users agree with sending their location due to personal privacy.

With new smartphone technology and low-cost air pollution sensors (Larkin and Hystad, 2017), a growing number of studies have evaluated exposure at a finer scale (Hwang and Lee, 2018; Liang et al., 2019; Sanchez et al., 2020; Steinle et al., 2015). Participants and study periods have ranged from five participants to 50, and from 12 days to 60, but from the reviewed studies, the number of participants was inversely proportional to the length of the study period. Steinle et al. (2015) and Sloan et al. (2016) validated $PM_{2.5}$ concentration with the monitoring station data to validate the experiment outcomes.

Compared to the spatial interpolation and dispersion modelling, the usage of wearable devices has benefited synchronising time and location and getting an accurate measurement of pollution by each microenvironment (ME). For example, Hwang and Lee (2018) collected GPS and PM_1, $PM_{2.5}$, and PM_{10} on a minute level basis and discovered that restaurants in Seoul had $PM_{2.5}$ at 96.1 ± 165.8 in summer and 85.4 ± 103.3 in winter, which was 5 times higher and 4.5 times higher than that of office spaces and residential parks. Although the selected studies used low-cost sensors, some studies (until the mid-2010s) had to pause the experiment to recharge batteries or to

[11]Locations were collected where the calls and messages were made

[12]"This model couples the land use regression model RIO, the road emissions model MIMOSA4 (taking into account COPERT4 emission functions, vehicle fleet and vehicle counts), and the Gaussian plume model IFDM." (Dewulf et al., 2016)

transfer files to the hard disk, but more recent studies were offered larger memory space, longer battery life, and even wifi and Bluetooth options to easily send files to a suitable device. Despite the finer spatial and temporal scale of pollution and location information that the GPS studies provided, the study period and the number of participants was not sufficient to generalise the findings because exposure levels can depend on travel modes, time and length of trips. Also, the trajectories and exposure patterns can be biased to a certain age group, a particular population subset, or people with particular travel patterns.

Smart cities can be another alternative to collect data of movement and pollution which can enhance the accuracy to evaluate pollution exposure at a finer spatial and temporal scale. Newcastle, for example, installed over 600 monitors (at lampposts) to collect pollution, weather, traffic (vehicle count, average speed, time to reach the next junction), soil, etc and transferred minute level data to the urban observatory with free access for the public. With a high density of pollution monitors across the city with high temporal resolutions, the database pairs nicely with the GPS data that will bring new research designs to facilitate evidence-based decisions for future urban air quality guidelines (Smith and Turner, 2019).

In summary, trajectory-based activities have provided a wider range of activity patterns, from OD matrices to GPS tracking. As people spend most of their time indoors or in transit, low-cost sensors to measure pollutants have improved the quality of exposure research by the time spent in various MEs rather than an estimated value from ambient pollution. However, biased profiles of participants and battery lifetime are still areas for improvement. Smart cities can substantially complement phone data by using smart sensors at every street that collects real-time data which can potentially offer more information between location and pollution exposure.

2.4.4 Summary of the Section

Section 2.4 has reviewed the existing literature of various methods that have been used for exposure assessment. Section 2.4.1 began with a uniformed interpolation that collected data from nearby stations, then reviewed studies that applied IDW and Kriging. While studies that used uniformed interpolation did not exceed the geographic boundary of a city scale[13], those that applied IDW and Kriging were applied from urban

[13]Except from Wong et al. (2004) who analysed the study before 2000

to nationwide areas. Section 2.4.2 reviewed dispersion models that considered vehicle emissions, meteorological effects, chemical transport, and even small-scale eddies that disperse vertically (more in the Appendix). This study reviewed a quasi-Gaussian model (ADMS) from the UK and an eulerian model (CMAQ) from the US. ADMS was the only model focusing on street canyon scale at 20m resolution, while the eulerian model took account of dispersion in a mesoscale at 1km resolution. In order to model a spatially high-resolution model to consider the variability of exposure across the city, the best solution was to combine ADMS for roadside dispersion and CMAQ for the background areas (Smith et al., 2016). Although both methods created maps to help indicate the impact on exposure levels, this study critiqued that creating maps with temporally rich but spatially poor resolution normally leads to a very smoothed map. Studies that use this method will neglect small-scale variability that might be detrimental for individual exposure assessment. Overlooking this granular dynamic will end up creating an ecological fallacy where 'all will have the same exposure level at this location', which is not true in reality.

More recently, GPS and mobile phone data have substantially benefited the problem of ecological fallacy by tracing individual participants' trajectories and their exposure levels. Considering zone-by-zone movements with OD matrices has helped to improve the population-movement to a fine-resolution, to get two snapshots of exposure between work (daytime) and home (after work/school). However, this method does not provide any information during travel. As mentioned in Section 2.3.1, the positional information and the time spent outdoors can diversify the exposure levels. Mobile phone data collected from cell towers provide positional information within the boundary, however, it depends on which mobile company the study chooses, and whether the people agree to share their data. GPS and low-cost sensors have also shown exposure at human height but the battery life and the limited number of participants can be a problem in conducting a long-term exposure study.

Even though personal exposure studies with light and portable devices provide accurate results, this can only be used for *ad hoc* analysis which is already proving useful but is perhaps least effective in comparing alternative routes. For example, if an individual has two options to arrive at their destination, which one is the shortest distance if the other has less traffic but takes a longer time to arrive. Current techniques have to send two people at the same time to get that result, but the decision can also be explored computationally with simulation techniques. One of the methods is

agent-based modelling (ABM), which can model pollution exposure at an individual level and test different pathfinding methods in a virtual environment. More details about pathfinding and exposure will be introduced in Chapter 5, but prior to that, the subsequent chapter takes a look at the basics of ABM and reviews air pollution studies that have used ABM.

2.5 Agent-Based Models in Air Pollution Studies

2.5.1 Concepts of ABM

ABM is a generic (or bottom-up) approach for simulating actions and interactions between agents to view their effects on the entire system (Gilbert, 2008). As mentioned, ABM produces macro-effects from micro-level interactions that construct a cornerstone of this complex methodological framework. Thus, within this paradigm, dynamic phenomena and evolution emerge when there is feedback between components (Castle and Crooks, 2006; Gilbert, 2008; Wooldridge, 2009). Also, *interactions* among autonomous agents can create new ideas which cannot be made by themselves. The main advantage is developing a model by inputting assumptions, simple or complex (equations), on active individuals to better understand how the system works (Auchincloss and Diez Roux, 2008).

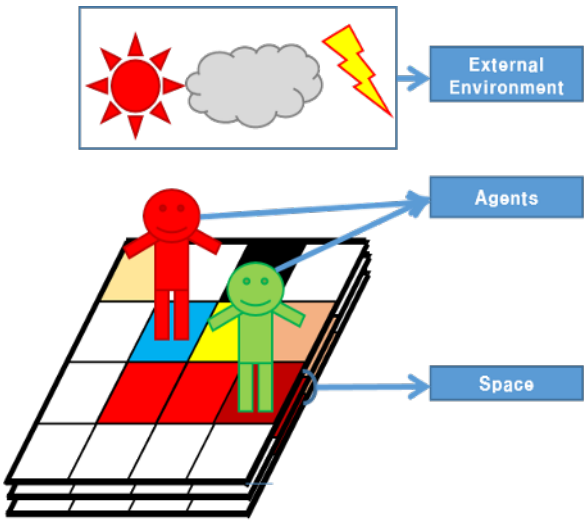


Fig. 2.8 Conceptual Structure of ABM

In the agent-based system, **agents** are defined entities which are heterogeneous, autonomous, and decision-making individuals (Gilbert, 2008; González, 2012). Agents do not necessarily need to be human. They could be animals, vehicles, plants, and so on, depending on the researcher's focus, but humans are generally used (Smith, 2012). Agents can be cooperative and adaptive, and their behaviour and movement are typically specified by a set of rules (Benenson and Torrens, 2004; Railsback and Grimm, 2011). The agents may also interact with nearer agents, based on their social network structure, or cluster into groups at different levels, such as households or administrative regions (Galea et al., 2010). Agents need not always have location information but can exist as a background parameter.

Space is a position where agents move and interact (see Figure 2.8). In many cases, ABM uses a group of discrete spatial grids that are spatially located and interconnected with each other through the neighbouring grids. In the tessellated space, the attributes of grids tend to be heterogeneous. Multiple structures of space including buildings, roads, and polluted sites, can influence and be influenced by the circumstances of the agents. There have been attempts to optimise model outcomes by tuning the grid resolutions or defining different types of neighbours e.g. von Neumann neighbourhood, Moore neighbourhood (Stanilov, 2012).

An **external environment** is a factor that can control agents and gridded space (González, 2012). This component is also defined as a parameter in software such as Netlogo, where once defined it affects any agents or spatial variables. Rainfall and weather, earthquakes, or other large-scale phenomena not directly affected by the agents could be considered as external environments.

2.5.2 ABM Model Environments and the Selection

Table 2.1 shows a number of software for agent-based modelling that have their own platform and syntax. These include AnyLogic (Weimer et al., 2016)[14], GAMA (Taillandier et al., 2010), MASON (Dunham, 2005), NetLogo (Wilensky and Rand, 2015), and RePast (Collier, 2003).

Amongst the software mentioned above, NetLogo was chosen for this study for a few reasons. Firstly, the Netlogo codes were relatively easier to comprehend. Every programming language has its own syntax, pros, and cons, but given the study design

[14]Free license for academic use

Table 2.1 List of ABM Software (partly adapted from Zheng et al. (2013))

Name	Scalability	Programming	License
AnyLogic	Large	Java	Commercial/free(academic)
GAMA	Large	Java	Free
MASON	Large	Java	Free
NetLogo	Intermediate	NetLogo	Free
RePast	Large	Java/C++	Free

and the learning curve, NetLogo was much convenient because agents are created through relatively well- structured syntax and provides a 'point and click' parameter settings in the interface. MASON, on the other hand, is faster and scalable but is more complicated to create both the agentset and GUI (Gilbert, 2008). Secondly, a vast majority of agent-based modellers are using NetLogo as their main software. This means that most of the queries can get answered through forums (e.g. comses.net, StackOverflow), journal articles, and example models. In fact, this study was inspired by a library model "Urban Suite - Air Pollution" (Felsen and Wilensky, 2007). Finally, although the speed of execution is slower than other software, R and Python developers have created useful packages to overcome the slowness by calling NetLogo in headless mode. R has RNetLogo (Thiele et al., 2014) and nlrx (Salecker et al., 2019), while Python has PyNetLogo (Jaxa-Rozen and Kwakkel, 2018). These packages can also take advantage of the automation process from parameter testing, data pre-processing and analytics, and visualisation using a few lines of codes; and the greatest benefit is that the users can launch the whole process on the cloud computing service, such as HPC (High-Performance Computing) or AWS (Amazon Web Services), for the sake of quicker execution and without exhausting a personal machine.

Indeed, embedded ABMs such as MESA in Python (Masad and Kazil, 2015) and R(Bauduin et al., 2019) might have been considered, but these packages were excluded for a couple of reasons. One was because the package arguments were appropriate to cover a list of basic examples, such as the wolf and sheep predator model (Wilensky and Reisman, 2006), but not able to simulate networks and GIS-based arguments that were crucial in Chapter 5. Although these packages may develop further in the future, they were still underdeveloped when the PhD project started.

Hereafter, the following sections may include some NetLogo arguments as an example to help the explanation.

2.5.3 Opportunities and Challenges in ABM

Opportunities

ABM has been spotlighted as a novel method in computational social science (CSS). The primary strength of ABM is the consideration of individual interactions and feedback loops over time and space (Auchincloss and Diez Roux, 2008; Gilbert, 2008; Koch and Carson, 2012; Smith, 2012).

Compared to deterministic methods, ABM enables heterogeneous agents moving and interacting across different locations without the need to aggregate up to population level, and to allow learning behaviour so that the structure of the feedback loops themselves can change over time. The dynamics of integrating agents and space have become significant in urban modelling, not only because it combines agents and space in a multi-scalar frame, but also it takes into account different representation of the landscape's heterogeneity (e.g. roads, junctions, parks, hills) during the agent's movement over time (Stanilov, 2012).

Second, the richness of the spatial and temporal resolution of ABM allows models to be made at many different scales (Wilensky and Rand, 2015). When a modeller designs an ABM from the beginning, s/he will initially simulate a proof-of-concept (or a demo) model that contains a small population in a small area. If modelling a sample of a population is valid, then this can be scaled to represent a larger population, although this is a complex task and depends on the model and the phenomena being explored.

Third, ABM can be used as a testbed to test hypotheses and to apply scenarios for future decisions, in other words, "What if..." scenarios (Stanilov, 2012). The models are not used as tools to predict what will happen but rather used to explain the behaviour of the system (Epstein, 2008; Stanilov, 2012). Scenarios include predictions of the stock market (Farmer and Foley, 2009) or health care policy (Gurram et al., 2019).

Finally, ABM can incorporate many different kinds of rule types and allows one to mix different types of calculations, making it a flexible tool able to represent multiple kinds of phenomena (An, 2012; Gilbert, 2008).

Challenges

Despite the opportunities, several challenges also exist that one should be aware of. First and most importantly, ABM is computationally expensive. The "price" can be broken down into three parts:

1. The difficulty of construction: ABM requires an individual settings (agent) and environmental settings (patches, though not necessarily), each with their unique attributes and behavioural rules (Crooks et al., 2018). This takes some time to pre-process the data and to call the dataset from the adequate package libraries (or extensions in NetLogo). For example, for the hourly PM_{10} from the monitoring stations one needs to first verify the proportion of missing data, fill in the missing parts after choosing a suitable imputation method, and then find ways to interpolate the data from the stations. Constructing individual behaviours is more complicated as decisions can vary by the heterogeneity of agents' characteristics, e.g. demographic features or mode of transport.

2. The computational power required for model implementation: The execution speed depends strongly on the computational power used as well as on the efficiency of the implementation. This statement can be controversial because the execution speed depends on the simulation environment, where more detailed results require more computational power (Wilensky and Rand, 2015). However, if the slowness of the model run is due to inefficient coding – for example, primitives of `with`, `max-n-of`, and `in-radius` in NetLogo – this can be solved, but if this is not the case, then it can vary by computer configurations and sometimes by the operating system[15] (Railsback et al., 2017).

3. Memory storage to post-analyse results: unaffordable memory is a challenge if many parameters are tested over many iterations[16]. Although the NetLogo helpdesk does not mention any fixed limit size, it has, by default, an upper limit of 1GB memory for storage with a 64-bit JVM (Java virtual machine) execution (Wilensky, 1999). Also, this thesis verified that NetLogo technically cannot create over 999 x 999 patches in the model; therefore, it is required to reduce either the spatial resolution or the spatial extent. In cases where a model

[15]The author tested that with the same condition Windows was the quickest among Windows, MacOS, and Ubuntu (Linux)

[16]Behaviourspace in NetLogo

consumes rapidly more RAM as it runs, then the modeller has to decide whether to simplify the model structure or to frequently export the file to the disk rather than keeping it. Otherwise, the system will run out of available memory and crash. High-Performance Computing (HPC) might solve the memory issue to some extent, but in this case, all post-analyses need to be operated through the RAM of the cloud, which is very time-consuming, i.e. waiting in the queue to interactively use remote HPC resources. Possible alternatives are to lighten the data by slicing files into several parts or save it to a file extension that compresses the data, e.g. feather[17] .

Another drawback can be the computational effort for testing 'free parameters' that are neglected in equation-based models (EBM). The term 'free' means the values are determined by the modeller (Wilensky and Rand, 2015). Compared to EBMs, ABM typically has extra parameters to take control of the model and calibrating multiple parameters with real-world data is time-consuming. As mentioned, stochastic effects between parameters can happen at every model run, and it gets more problematic when the model has too many parameters to handle or the model testing gets very computationally intensive. Thus, each parameter to be tested may need multiple models runs to deal with the effects of stochastic variability.

Furthermore, ABM, as with all social system models, suffers to validate ground truth data even if there is enough data. ABM can be aggregated to match census tract data and then calibrated; however, one may need more data as a result. But in any case, even if individual data were available, the model would not match it because of errors and stochastic variability. Thus, the ABM's approximation would only be one way to calibrate the model.

2.5.4 ABM Studies in Relation to Air Pollution Exposure

ABM has been used in environmental public health research since the early 2000s (Auchincloss and Diez Roux, 2008; Holland, 2003). Both studies mentioned here agreed with the significance of individual interaction models when discussing place effects on health. Auchincloss and Diez Roux (2008) realised that traditional regression approaches can only discover the association between independent and dependent

[17]Feather is a fast, lightweight, language agnostic and easy-to-use binary file format for storing data frames

variables but overlook the processes of dynamic interactions between heterogeneous individuals. The dynamic features of ABM, therefore, have been applied to both communicable or non-communicable diseases in which individual behaviour can cause a population-level change (Galea et al., 2010). With ABM, however, simple but random elements might be applied to individual actions, allowing one to ascertain measures of uncertainty.

In general, ABM research has been conducted in traffic simulation and emission studies using MATSim[18] (Hülsmann et al., 2014, 2011) or SUMO[19] (Lemos and Pasin, 2016), or pathfinding algorithms to better suggest transport routing in large cities (Crooks et al., 2019; Manley and Cheng, 2018; Manley et al., 2015a,b). However, only a few have combined the two topics. David and Don (2012) developed an urban pollution model (EPICast-API) that built a hybrid model of ABM, an air dispersion model (TAPM), and a dose-response model[20] to estimate possibilities of mortality and morbidity in the greater metropolitan area of Sydney. In more detail, the model generated human movement and estimated its pollutant dosage by their time of exposure. The authors validated the result of respiratory hospital admissions and emergency room visits. As far as this project is aware, this was the first approach using ABM that used a synthetic population rather than that of sampled to investigate pollution vulnerability. However, exploring the relationship between mortality statistics at a coarse grid scale ($1.5km^2$), and only for a single year may miss the effects of spatial variability and cumulative exposure. In addition, the model did not break down their results by social class.

Another hybrid model was published by Gurram et al. (2019), where the authors evaluated the unequal exposure to NO_2 in the population of Tampa, Florida using high-resolution data. The model integrated: 1) an activity-based travel demand model for the population to schedule the timelines in a day (DaySim, individual-scale), 2) a microscopic traffic assignment model that can simulate thousands of cars at a city-wide domain (MATSim, individual-scale), 3) a mobile source emission model to generate NO_x from the tailpipes (MOVES, road-scale), and 4) a pollution dispersion model to

[18]Multi Agent Transport Simulation

[19]Simulation of Urban MObility

[20]Dose-response model is a statistical approach to measure personal health associated with pollution exposure levels. Dose is defined as an amount of agent –"a chemical, biological or physical entity that contacts a target"– that enters a target –"any physical biological or ecological object exposed to an agent"– in a specific time duration after crossing the absorption barrier, and response is a biological response to an agent: different function, morbidity, or mortality (Ott et al., 2006)

disperse NO_x when the pollutants are released from the vehicle (R-LINE, road-scale). The integrated model gave great insights that the models, which have different scales and are already complex on their own, brought together to assess individual exposures according to their time-activity patterns. However, this model overestimates the results by associating the likelihood of health risks with lower income, non-whites, a deprived neighbourhood in Tampa city, having only executed the model for a single day. Even if short-term effects of NO_x can trigger immediate harm to the human system, it will be very unlikely that people living in a deprived neighbourhood or who are not well-off would be regarded as more at risk without knowing their underlying disease histories. Another limitation is that the study only chose passenger cars according to its connection with other models, but in the real world, long-distance commuters and logistics can contribute more pollution load to the atmosphere.

Most recently, researchers King's College London and other researchers around the UK have initiated an Air Pollution Exposure model (APEx) to suggest ways of protecting the most vulnerable groups from air pollution . The proposed model will take into account the urban models of air pollution, buildings and urban form, and create an agent-based model to simulate the human behaviour of a sampled population. To date, no papers have been published yet, but considering that they already have a solid urban air pollution model - the London Hybrid Exposure Model (LHEM) – up and running (Smith et al., 2016), the only remaining job might be to examine the movement patterns of agents, particularly the vulnerable populations they are targeting, and apply current or augmented air policy scenarios to envisage possible improvements of health outcomes.

2.5.5 Summary of the Section

In summary, in the agent-based modelling section, most of the grounds were devoted to exploring concepts and important features of ABM. One reason was to simply understand the fundamental knowledge of complexity theory and seek studies that are relevant to this project. The other reason was to observe the possibilities of modelling health vulnerability and exposure with a synthetic population by tracking agent trajectories in a heterogeneous space. While spatial interpolation and dispersion methods have worked closely with air pollution and health, relatively fewer research studies have been done using agent-based modelling. This might be because quantifying a health response to pollution exposure requires a lot of assumptions and conditions,

and more risks in long-term modelling might assume more uncertainty in terms of how to parameterise a person's ageing and the structural dynamics of urban growth. Despite a few reviews, both studies have selected their study environments in an urban context (i.e. Sydney and Tampa), however, the spatial and temporal scales were very different.

The Sydney model investigated the mortality and morbidity of the population based on an agent-based simulation and a pollution dispersion model (TAPM). However, there were some gaps in oversimplifying the results derived from a coarse resolution ($1km^2$), and only estimating risks of a synthetic population, neglecting the degree of difference in social class as well as their trajectories.

The Tampa model was established to assess the near-realistic exposure to NO_2, by combining an activity model for the whole population of Tampa, a microscopic transport model that transported individual Tampa citizens, an emission model that generated NO_x, and a dispersion model that dispersed NO_x to the closer atmosphere. Despite the integration which allowed the study to estimate individual exposure with more accuracy, the simulation period was too short to evaluate whether exposure levels are likely to be unequal between different social groups. A new air pollution exposure (APEx) simulation for London is currently under development. The project is promising not only because it has a high-resolution air pollution model completed, but also because they have unique personal exposure measurements gathered from several research campaigns which can be used for calibration. This will all come together in a large-scale ABM that will show how people are exposed to air pollution according to their patterns and use scenarios to help mitigate health risk by the augmented clean air policy.

2.6 Chapter Summary

- The study has looked though different types of pollutants that can harm human health.
- This study chose PM_{10} and NO_2 for the empirical chapters.
- This study contextualised the association between pollution, exposure, and health risk, and how sensitivity and adaptation can change the patterns of exposure.

- Exposure studies are categorised into population exposure and individual exposure. And this normally is tied with the method used. This study reviewed spatial interpolation and dispersion models, and criticised that using a static population does not fully account for the temporal variability that individuals might experience; using spatially poor station levels too does not consider the small-scale variability that can happen in the real world.
- OD matrices have filled the gaps of activity patterns to some extent but have not been able to consider the routes to which the individuals take every day. Mobile phone data cannot trace individual data due to privacy issues, and the representation of users strongly depends on the mobile company within the country. Low-cost sensors and GPS signals have shown minute based trajectories and pollution levels at human height; however, these have shortness of battery problems and a number of participants to test different routes.
- ABMs can be an alternative to the above methods as they can evaluate population exposure for all heterogeneous agents, conduct short or long-term simulation studies that can take up to a decade, assign cognitive decisions for an individual's pathfinding, and experiment with scenarios that might happen in the future. However, this too is a model that takes into account many assumptions, and constructing an ABM taking into account all the individual attributes takes a long time until completion (approximately two years). With huge output files, it might also need to be subsetted and restructured to produce the results with the computer memory secured.
- Published studies of Sydney have introduced a pollution dispersion model and a dose-response model to estimate the health risk of air pollution. Using a static population and a coarse resolution oversimplified the result. A more recent study in Tampa used a microscopic simulation method that considered traffic, pollution and population activity to assess the uneven distribution of NO_2 exposure across the city. However, the execution period was too short (one day) to determine the effect of socioeconomic factors.

Chapter 3

A Comparison of Spatial Interpolation Methods for the Estimation of Seoul's Air Pollution: A Support Tool for Population Exposure Modelling

Chapter Overview

Understanding the association between pollution exposure and the deleterious effect on population health is a vital precursor for modelling. Chapter 3 takes a step forward from the previous chapter, by exploring the spatial and temporal aspects of air pollution by employing geostatistical methods. This chapter initially argues that the comparison amongst methods should not merely conclude with error measurements but should add more information to the difference; secondly, it argues that a finer temporal scale is needed in pollution studies due to human mobility and daily and seasonal dynamics of pollution; and thirdly, compared to dispersion models, spatial interpolation methods in air pollution still have a great advantage that is faster, open, and free to use. This chapter aims to compare spatial interpolation methods that can support population exposure to daily air pollution. Seoul was selected as a case study.

This chapter firstly explores the pollution data in different time sequences and types of monitoring stations. Then it compares two spatial interpolation methods, Universal Kriging, and GAM (Generalised Additive Model), with additional weights on road layers as an effect of roadside pollution fields. After the calibration and validation, this chapter finally discusses issues relevant to interpolation methods and health effects.

As a result, a remarkable contrast was found between roadside and background areas of NO_2 on a daily basis: this allowed the enhanced levels of NO_2 on roads to be mapped. PM_{10} had a small variance in daily cycles but had greater seasonal oscillation. This means that the population may be exposed to high levels of PM_{10} in particular seasons regardless of location while that of NO_2 might be more variable. Neither of the interpolation methods was noticeably superior to the other, but the sparse station data meant that only very smoothed large-scale fields could be recovered, which did not accurately represent the extremes observed at individual stations.

Keywords: Seoul, Spatial Interpolation, Kriging, generalised additive model (GAM), Road-weighting, road non-road ratio

3.1 Background

Population health has been seriously threatened by daily ambient air pollution in South Korea. Despite the efforts to legislate national pollution standards, daily peaks in recent winters and springs have already exceeded the standards countless times. During 5th-8th March 2019, the entire country experienced over 200µg/m^3 of PM_{10} and smog episodes; the national authorities warned everyone to reduce outdoor activities. As most of Seoul's areas tend to experience disastrous levels of pollution frequently, residents can be exposed to air pollution unconsciously, which in the long-term can lead to respiratory or cardiovascular ailments (Zhang and Batterman, 2013). Thus, understanding the spatial and temporal aspects of air pollution and its relationship with exposure is crucial.

Exposure research has exploited spatial interpolation (SI) to investigate the relationship between ambient air pollution and population exposure in a spatial context – "which places have high air pollution?" and "how many people can potentially become unwell from high episodes?". SI is a statistical method that can compute pollution fields over a wide area with a given set of point measurements. SI can mainly be split into

a group that follows the assumption of spatial autocorrelation (e.g. Inverse Distance Weighted (IDW), Kriging), and a group on statistical inference e.g. generalised linear models and generalised additive models (Wood, 2019). Methods that take spatial autocorrelation into account assume that the values tend to be more similar when closer together, whereas methods that use (spatial) statistical inference delineate the inferential surfaces over a region by minimising the residuals of the model.

However, when air pollution monitoring stations such as those in Seoul are small in number compared to the size of the city, the estimation of the potential population at risk due to air pollution can be completely different depending on the measurement method used (Wu et al., 2019). Wong et al. (2004) used four spatial interpolation methods – spatial averaging, nearest neighbour, IDW, and Kriging - to estimate children's exposure to air pollution across the USA. The outcomes of the four methods only showed a small difference of PM_{10} and O_3 where the monitoring stations were denser, for example in the northeastern cities and urban California, but was difficult to predict in the mountainous regions and high-altitude zones. Aalto et al. (2013) compared a Generalised Additive Model (GAM) and an Ordinary Kriging (OK) to predict monthly mean temperature and precipitation and found that the GAM outperformed other methods by a small amount but the biased distribution of stations ("concentrated in the urban areas" of Finland) might have evened out the RMSE measures.

In addition, previous studies have provided tentative estimates of population risk based on annual or monthly statistics, but the aggregated figure lacks the potential for including acute injuries after an abrupt pollution rise, which may be more severe. However, there is likely to be a greater difference when the population at risk is measured at a finer temporal scale. This might support guidelines for surveillance of short-term exposure.

This chapter aims to compare spatial interpolation methods that can support estimating population exposure to daily air pollution in Seoul. This study compared Universal kriging (UK), Generalised additive model (GAM), UK with additional road effect, and GAM with additional road effect to model PM_{10}, and NO_2 in Seoul as an intermediate phase of pollution modelling. Compared to previous studies, this chapter generates outcomes on a 12-hour basis to understand the daily cycle of the pollution over the city and superimposes road effects to take into account small scale variables that might be neglected in the typical spatial interpolation outcomes.

3.2 Spatial Interpolation for Pollution Review: A Selection of GAM and Kriging

In pollution and related studies, there have been some comparisons of SI methods (Wong et al., 2004; Zhang and Shen, 2015). Wong et al. (2004) used IDW (Inverse Distance Weighting), Kriging, and Nearest Neighbour methods to predict O_3 and PM_{10} in the US where the monitoring stations are unequally distributed. They discovered that the predicted areas, which seldom had stations nearby, had similar values from different methods, and the landscape of California - flat and static landcover, with numerous monitoring stations - was ideally the best state to conduct this study. This study used leave-one-out cross-validations (LOOCV) to estimate the accuracy between the model and the observations with R^2, but noted it was difficult to get reasonable evidence of spatial autocorrelation for the entire country as the spatial continuity of the semivariogram that appeared well in the northeastern and southwestern regions that the others due to the spatial variability of station installations. Zhang and Shen (2015) compared IDW, Ordinary Kriging, and Trend Surfaces to predict $PM_{2.5}$ in Xian, China, using 13 monitoring sites. The accuracy was compared by using Mean Error (ME), Mean Absolute Error (MAE), Mean Relative Error (MRE), Root Mean Square Error (RMSE) and the System Error (SE). According to the indices, IDW was the most accurate tool. Deligiorgi and Philippopoulos (2011) used IDW, Kriging, Nearest Neighbour, and Artificial Neural Network (ANN) to predict NO_2 and O_3 in Athens, Greece. The accuracy was cross-validated with the leave-one-out method, resulting in MAE and MAPE (mean absolute percentage error) values. The study revealed that ANN outperformed the other methods both on NO_2 and O_3, but other methods were more accurate at some stations.

To the author's knowledge, there have so far only been a few publications that compared Kriging and GAM (Aalto et al., 2013; Parmentier et al., 2014). Aalto et al. (2013) imported precipitation and temperature data from stations over Finland for response variables, while elevation, sea percentage, and lake percentage were used to support spatial variability in the model. The study used RMSE and MAD (Mean Absolute Difference) to evaluate the results. GAM was revealed to be the best overall method for predicting the monthly mean with only a very small difference, but a wide variation was observed through seasons. For precipitation prediction, Kriging with external drift (KED) and residual Kriging (GK) performed better. Parmentier et al.

(2014) compared GAM with UK, and geographically weighted regression (GWR) to predict 1 year of monthly mean temperatures in 2010. This study also concluded that GAM outperformed the other methods but the difference was small, mainly because the 1km spatial resolution was too large to predict the spatial variability at small scales.

Apart from Wong et al. (2004), all of the studies used statistical measures to evaluate the accuracy. However, it is difficult to determine an appropriate method (no methodology is superior over others) since the results vary by spatial extent or data density. In other words, the predictions can vary by the number of stations and their evenness of locations, or some methods can work better in one situation but worse in another compared to other methods. This study also argues that cross-validation statistics (e.g. RMSE) should be reconsidered when applying spatial interpolation. This is because the model accuracy is dependent on the model setup or the backbone formula – such as variograms, even if the predicted outcome is similar. This will be discussed in the Discussion and Conclusion section.

In this respect, using spatial interpolation for urban air pollution is very challenging as the concentration will not only differ by the time of day but also in the areas which people visit. However, given that many cities still have a limited number of monitoring stations - also sparsely distributed - and that recent research still uses this approach, it is worth examining the pros and cons in pollution studies. This will be summarised in the method section.

3.3 Methodology

This chapter comprises three stages to achieve the research goal. The first stage is to tidy and explore a time-series of pollution around Seoul. The cleaning job from this stage links to the next stage of comparing interpolation methods. This study employs Universal Kriging (UK) and GAM (Generalised Additive Model) to draw estimated maps of PM_{10} and NO_2 in Seoul. In `R 4.0.2`, data cleaning and exploration were implemented using the `tidyverse` meta-package (Wickham, 2019), Kriging interpolation with `gstat` (Pebesma, 2019), and GAM interpolation with `mgcv` (Wood, 2017).

3.3.1 Data Collection

Pollution

This chapter explores the spatial and temporal patterns of pollution in Seoul as well as the extent to which background and roadside stations are similar or different. The NIER (National Institute of Environmental Research) provided 6 pollutants, PM_{10}, $PM_{2.5}$, NO_2, CO, SO_2, O_3, which were released as an hourly aggregation. This study selected PM_{10} as the main source that can causally result in human cancer and NO_2 as the main source that harms respiratory symptoms, whether constantly or instantaneously (Europe, 2013).

Pollution data were collected from two different types of stations: background stations (installed on the rooftops of district offices), and roadside stations (close to the major roads which are strongly affected by nearby traffic). This study took into account 57 urban background stations and 19 roadside stations that were within 10km from the city boundary (see Figure 3.1). Amongst these 76 stations, Seoul itself had 25 background stations and 15 roadside stations. Roadside data were retrieved from the 19 stations installed on the roads near the city centre, 8-lane junctions, and a highway entrance. The possible download period was between 01/01/2010 01:00 and 01/01/2018 00:00. Units are measured in ppb (parts per billion) for NO_2, and $\mu g/m^3$ for PM_{10}.

Road Layout

Road layout was provided by the Korean Transport Database (KTDB). Seoul had 59,319 road segments in total, and the attributes of each segment included node ID, link ID, number of lanes, maximum speed, traffic light density, and road ranks. In Table 3.1, Road segments were grouped by their road ranks as a level to distinguish different pollution loading. Rural and county roads 105, 106, and 107 were excluded due to a small fraction existing in Seoul. The roads were regrouped as general roads (103, 104), and highways (101, 102, 108) to distinguish the pollution loads by two road types.

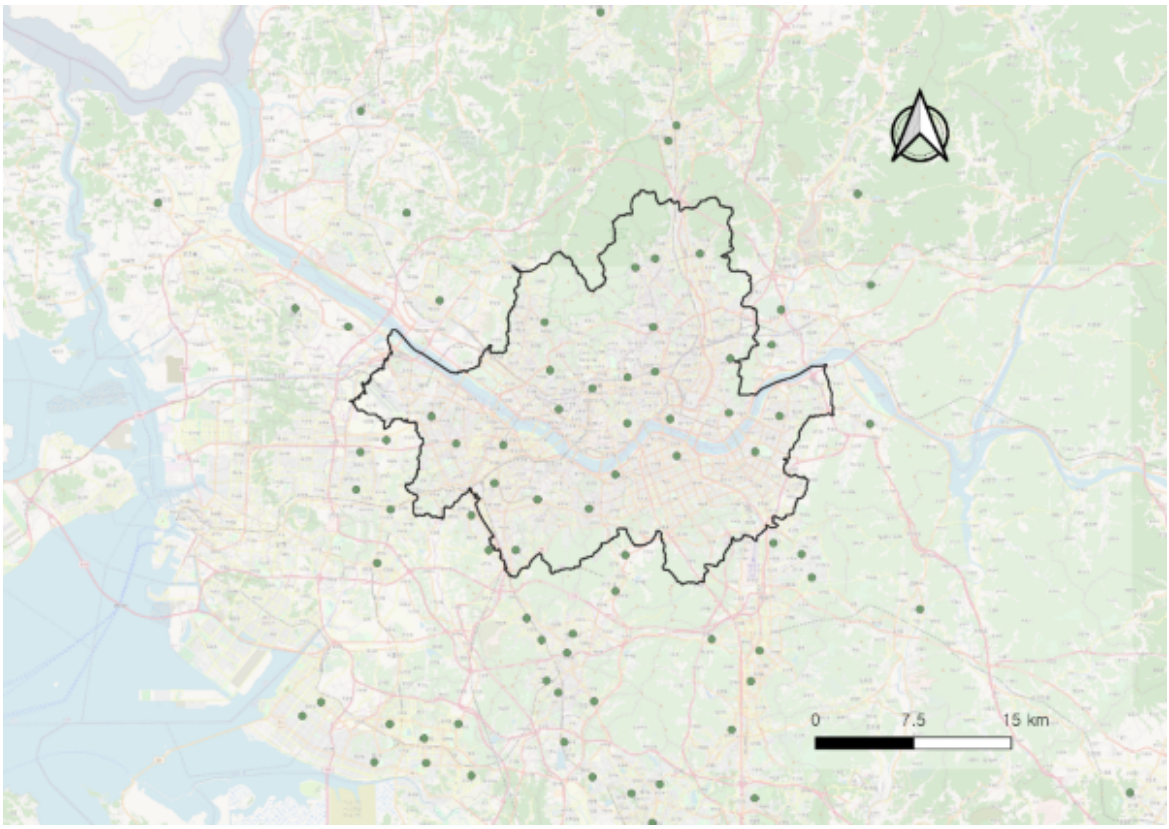


Fig. 3.1 57 Background pollution stations that are considered for spatial interpolation (boundary area: Seoul)

Table 3.1 Road hierarchies and groups

Rank	Name	Group
101	Highway	Highway
102	Urban highway	Highway
103	National road	General
104	Metropolitan road	General
105	Rural road, government-supported	-
106	Rural road	-
107	County road	-
108	Highway ramp	General

Table 3.2 Summary statistics of NO_2 and PM_{10} from 57 Seoul background stations

Type		Season	Mean	Min	Max
Mean	NO_2	Spring	34.6	16.9	64.2
		Summer	25.4	9	52
		Autumn	31.5	10.7	60.2
		Winter	37.5	13.6	67.3
Mean	PM_{10}	Spring	62.1	28.2	181.7
		Summer	39.4	15.2	126.8
		Autumn	40.8	16.9	199.6
		Winter	56.2	21.3	161.7
Max	NO_2	Spring	124.8	79	274
		Summer	122.4	53.6	255
		Autumn	118.5	69	340
		Winter	122.1	70	240
Max	PM_{10}	Spring	235.1	105	1160
		Summer	146.6	63.7	334
		Autumn	172.4	80	1613
		Winter	205.4	62	1086

3.3.2 Data Exploration

This section examines the spatial patterns and associated temporal patterns of NO_2 and PM_{10} between 2010 and 2017. Using the raw data of hourly pollution from the Seoul Institute, the initial analysis is to explore the temporal time series of pollutants from 76 different pollution stations across Seoul (57 background stations, 19 roadside/urban-highway stations). The same data were used to compare background and roadside concentrations.

Weekly averaged data were used for the temporal graph of pollution by years to understand the overall seasonal and interannual variability of pollutants and partly to remove outliers. The following exercise used a boxplot of weekly means by 24-hour bins to identify the daily variation of pollutants.

Exploring Daily Air Pollution in Seoul Stations: Time-series Analysis

Figure 3.2 shows a time series of NO_2 and PM_{10} across 57 background stations around Seoul from 2010 to 2017. Each point in Figure 3.2(a) represents a weekly average while

Figure 3.2(b) represents a weekly maximum value. The weekly average was used to present outcomes with a less cluttered view but this study is aware that the extreme is smoothed out. Each graph indicates a temporal oscillation by season, but the peak seasons differ by pollutants. Seasonal averages are presented in Table 3.2.

Overall, both pollutants seemed to have indicated a downward trend over time (see Figure 3.2). Although the pollution charts have different y-axes some points of comparison can be made NO_2 showed a wide range in daily oscillations from traffic or other combustion processes while PM_{10} tended to be similar between weeks but showed a wider range between seasons (17-200µg/m^3). The maximum NO_2 depicted a mean close to 100ppb which is 50% higher than the nation's daily standard of 60ppb. The maximum PM_{10} maintained a mean of around 180µg/m^3 throughout the whole period and often skyrocketed over 1,000µg/m^3. Thus, the time series of NO_2 can be summarised as frequent oscillations between days (according to Figure 3.4) and seasons and PM_{10} as seasonally oscillating but with occasional very high episodes.

Throughout the whole period, summer was distinctively low for both pollutants, where the average NO_2 was at 25.4ppb and PM_{10} at 39.4µg/m^3. These were around 12ppb and 23µg/m^3 lower than the highest season. As with the results from Figure 3.2, the highest mean NO_2 for all seasons except winter managed to fall under the daily standard, but the mean in the maximum NO_2 were twice as high as the standard. The highest PM_{10} of the mean values was all over 100µg/m^3 and 3-4 times higher than each seasonal average. The average[3] maximum of PM_{10} in spring was 235.1µg/m^3 which was almost twice as high as summer. Peaks from both pollutants tended to appear during the transition period of autumn-winter and winter-spring when the air mass gets drier and transnational air pollution from the continent arrives in Korea.

To understand the daily cycles of pollution, the dataset was illustrated as a boxplot (see Figure 3.4). Overall, PM_{10} (51.77µg/m^3) remained near the annual average where small rises were monitored during 09h-14h but the values were maintained during the rest of the day. Unlike the mean time series, the maximum PM_{10} significantly exceeds national standards regardless of the time of the day. This implies that despite the lesser daily variability shown for PM_{10}, several occurrences of pollution episodes can cause peaks that surge up to 1,500µg/m^3. On the other hand, NO_2 had two peaks during 07-10h and 22-24h. The initial rise was most likely due to the traffic in rush hours. The second rise was discovered between 22-02h, although another long rush hour took place between 18-20h, perhaps because the residual layers of substances such as ozone

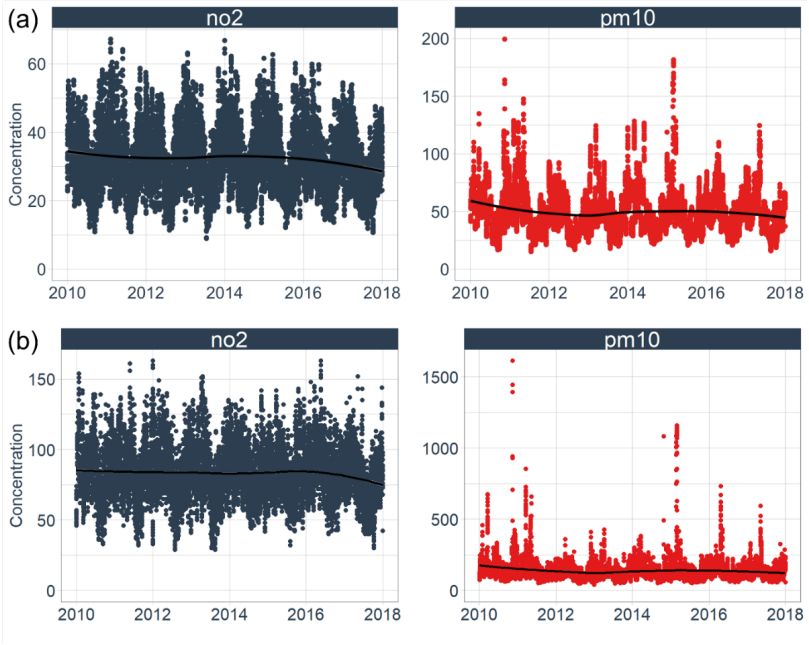


Fig. 3.2 Time series of NO_2(ppb) and PM_{10}(µg/m^3) from 57 Seoul background stations between 2010 and 2017 by weekly (a) mean and (b) maximum concentrations

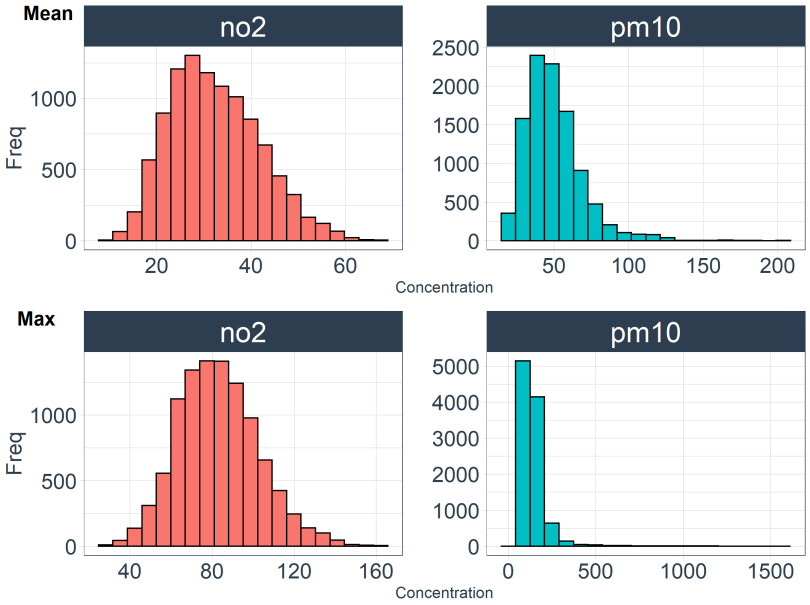


Fig. 3.3 Averages of NO_2 and PM_{10} from Figure 3.2 converted to histograms

remained at nighttime and the chemical decay in the absence of sunlight might have caused additional oxidation to NO_x, whereas reduction in NO_x due to photochemistry would be active during the day (Kim et al., 2016). From the daily plots, the occurrence of high values of NO_2 did not vary much by seasons but varied in hours of the day.

Comparison between the background and roadside stations

During 2010 to 2017, the annual average for roadside stations was seemingly higher than that of background stations in NO_2 and was higher for PM_{10} (see Figure 3.5(a)). For NO_2, the declining trend was observed in all seasons, however, winter had the smallest difference between road and background emissions that were at 38ppb and 55ppb. Although summer had the lowest background emissions at 24ppb, the roadside average was 42ppb. However, the maximum values of the background stations showed a considerably high concentration at roughly 113ppb, 100ppb, 104ppb, and 112ppb in spring, summer, autumn, and winter with the highest at 163ppb, 143ppb, 143ppb, and 163ppb respectively (see Figure 3.5(b)). By contrast, the road concentration was at least 20ppb higher than the background, at 135ppb, 144ppb, 132ppb, and 131ppb in order of season, but peaked at 274ppb, 255ppb, 340ppb and 240ppb respectively.

The differences between background and roadside stations in NO_2 is likely to be contributed to by traffic emissions which normally dilutes with the photochemical reaction of ozone within a few hours (Oanh, 2012); however, the distribution of the maximum values implies that people whose metabolisms are vulnerable to sudden air pollution rises may suffer from a deterioration of their health from only short-term exposure.

By contrast, PM_{10} showed less difference between road and background and seasons (see Figure 3.5(a)). The difference was less than 10µg/m^3 on average, which presumably implies no difference in some places. PM_{10} in spring appeared higher than in the rest of the regions which is speculated to be the transboundary air pollution consistently blowing from the Eurasian continent (Shapiro, 2018). Figure 3.5(b) illustrates the seasonal averages of the weekly maximums for each year, which were at 244µg/m^3, 151µg/m^3, 177µg/m^3, and 212µg/m^3 in spring, summer, autumn, and winter at background stations, while roadside stations had 226µg/m^3, 141µg/m^3, 166µg/m^3, and 198µg/m^3 in seasonal order.

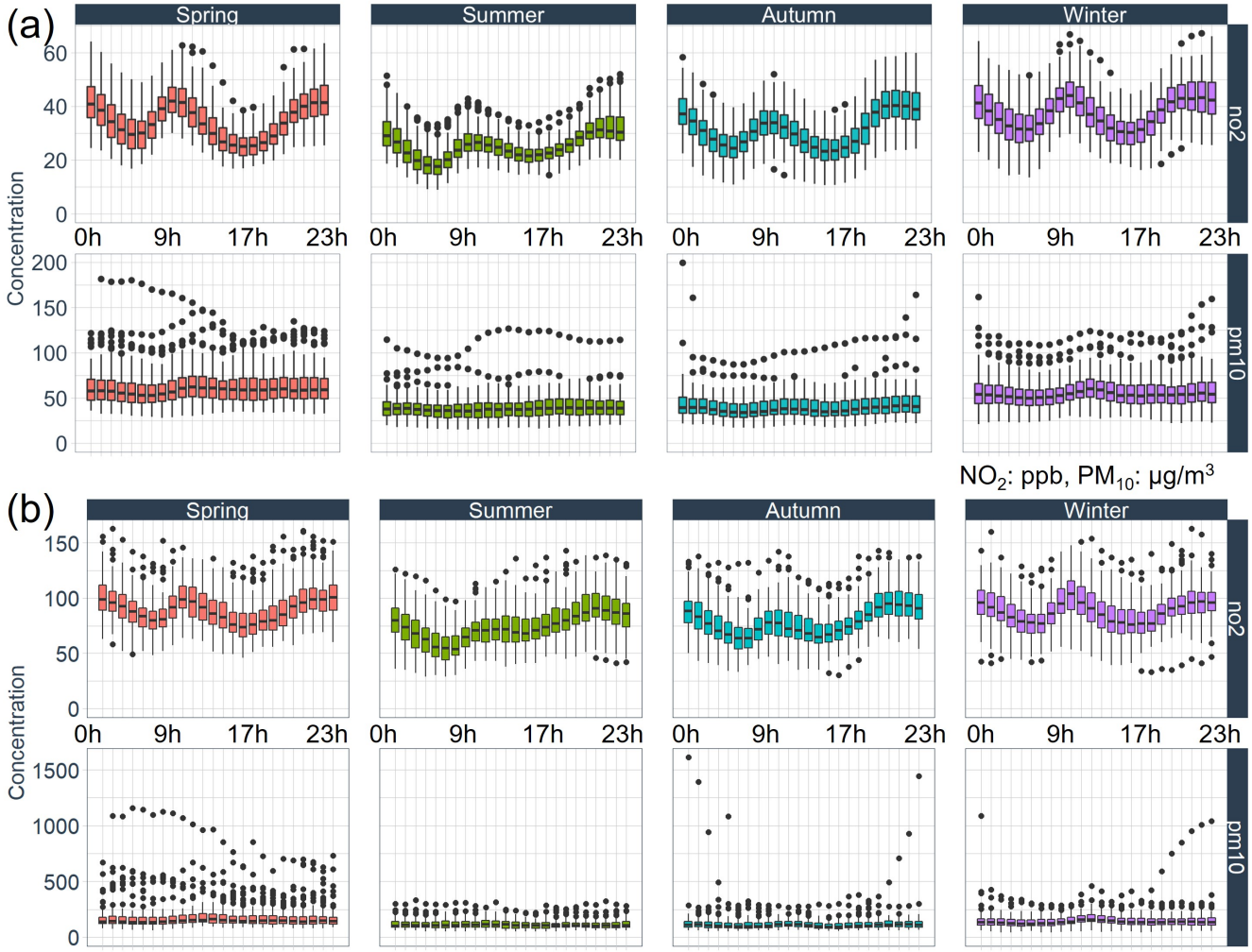


Fig. 3.4 Figure (a) illustrates the hourly mean of NO_2 (top) and PM_{10} (bottom) from 57 monitoring stations around Seoul collected between 2010-2017. NO_2 shows a wider variation within a day with two peaks occurring at 07:00-09:00 and 22:00-24:00. By contrast, the trend of PM_{10} turns out to be quite plateaued throughout the day, but most of the hours have outliers that exceed 100μg/m^3, meaning people are likely to be exposed to extreme figures of particulates in some parts of Seoul. As a reference, Figure (b) illustrates the hourly maximum values of NO_2 (top) and PM_{10} (bottom) from 57 monitoring stations around Seoul collected between 2010-2017. NO_2 has a similar temporal pattern compared to that of mean but with the concentrations distributed across 100ppb, while the maximum values of hourly PM_{10} have skyrocketed to 1,500μg/m^3

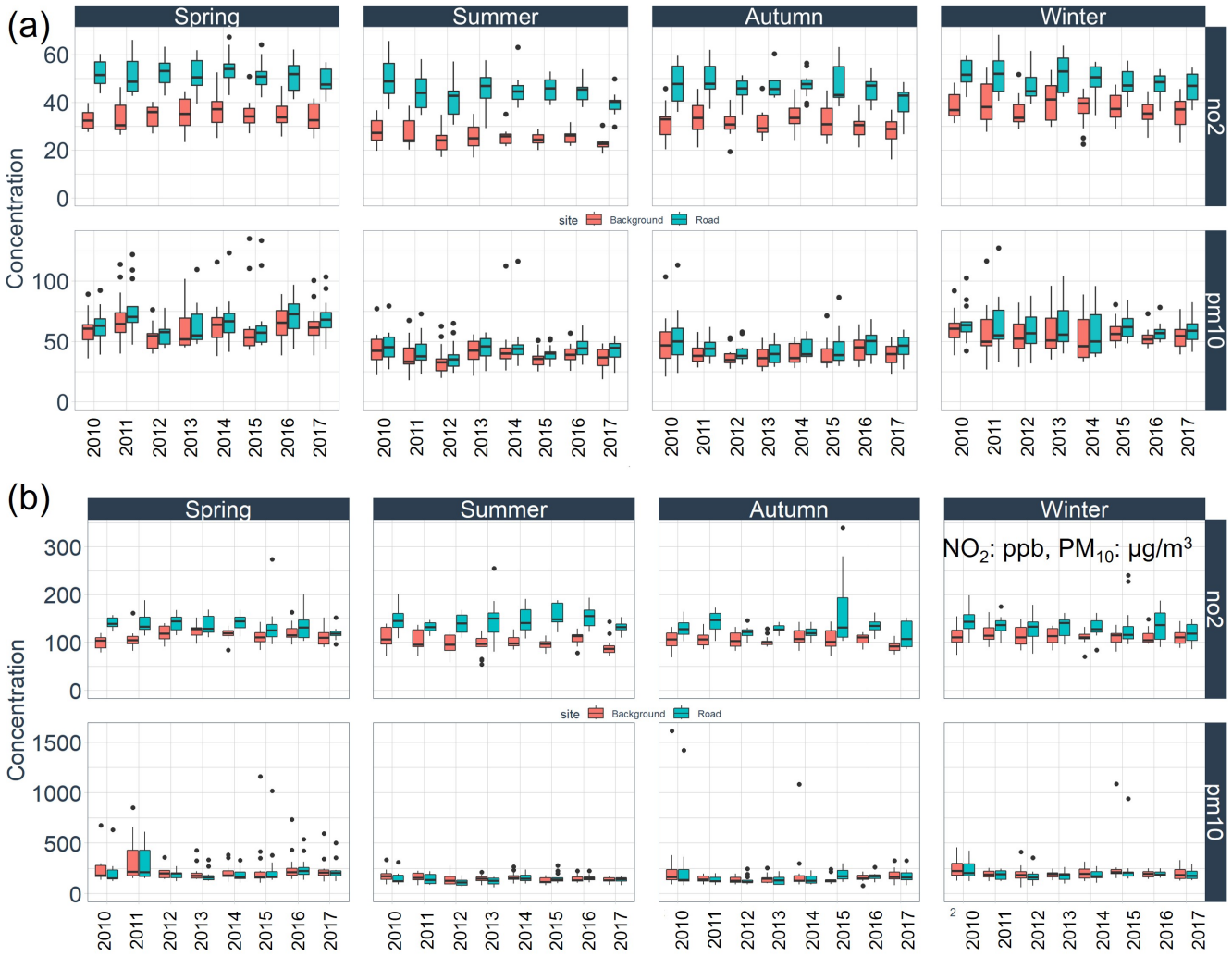


Fig. 3.5 A comparison of background and roadside stations for two pollutants in four seasons of 2010-2017 by selecting the (a) mean and (b) maximum. The units for NO_2 are measured by ppb (particles per billion), and µg/m^3 for PM_{10}.

3.3.3 Kriging: A Geospatial Interpolation Method

Kriging is a geostatistical method that interpolates an unknown location using the characterized mean and variance structures (Kim et al., 2014; Kumar, 2007). It assumes that the overall mean, distance, and variance of all observations are spatially autocorrelated from Tobler's First Law of Geography: *"Everything is related to everything else, but nearer things are more related than distant things"*. For example, pollution concentrations less than 1 metre apart are likely to be similar, but less likely when the distance becomes larger. This context of the variability between the two values and their grouped distances (e.g. 0-100m, 101-200m) can be structurally applied over the region with a conceptual semivariogram. Once the semivariogram is produced, then the kriged map is produced together with an error map.

To create a kriged map, there are a few steps to follow: 1) modelling an empirical semivariogram, 2) fitting the empirical model with a mathematical model, and 3) choosing the type of kriging according to the fit of the assumptions (i.e. data normality, stationarity, and whether the data has trends).

Understanding Semivariograms

Spatial autocorrelation (or spatial dependency) is assessed by a semivariogram. Semivariograms apply the squared differences of measurements against distances between pairs of data points (see Equation 3.1). This means that all the pollution values at the 57 stations are compared between one another, and once calculated, the semivariogram is drawn. The conceptual semivariance can be estimated as below.

$$\gamma(h) = \frac{1}{2N}(h) \sum_{i=1}^{N(h)} Z(s_i) - Z(s_\mathrm{i} + h)^2 \tag{3.1}$$

- N(h)= number of pairs of observation points with distance `h`
- Z = field that holds a height or magnitude value for each point (z-values)
- Z(s_i)and Z(Si+h) = sample data pairs at a distance `h` (Luo et al., 2008)

Directions are not considered in the current formula; however, this can be taken into account depending on the aspect of the study, for example, if the wind direction

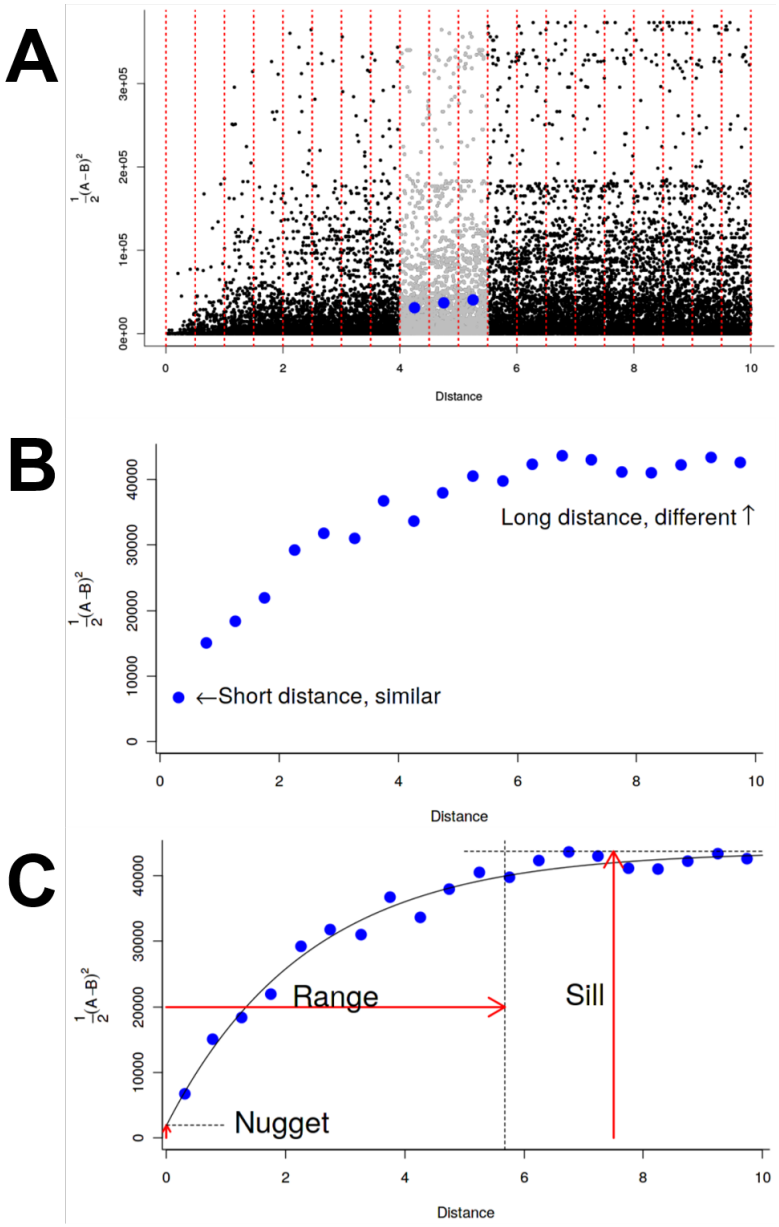


Fig. 3.6 Conceptual process of kriging measurements (captured from Datacamp.com lectures)

is dominantly affecting the pollution concentrations. Further information regarding anisotropy is mentioned in Section 3.3.5.

Figure 3.6 explains the process to generate a semivariogram. First, the distance against half-squared distances between all pair points is plotted in the variogram cloud (see Figure 3.6a). The dots in the variogram cloud are allocated to an arbitrary number

of imaginary bins (10 bins are used for this study). The mean of each bin becomes a representative point which normally has an upward trend as the distance increases to some extent and then levels off (see Figure 3.6b). This means that the closer the distance, the more similar the values are, and vice versa. Finally, an asymptotic curve, termed semivariogram, is drawn through the points (see Figure 3.6c). Range[1], partial sill[2], and nugget[3] are the key parameters inside the model, and the kriged map is drawn based on the semivariogram (Law and Collins, 2019).

Why is the Semivariogram Difficult to Model?

Semivariograms are extremely sensitive in terms of their fit with empirical variograms and this hasn't been discussed thoroughly in the literature. If the points perfectly fit into a semivariogram model, the observed and the estimated value should be identical and there is no problem in creating maps (Aalto et al., 2013; Luo et al., 2008).

Here is a list of reasons why variogram fitting is difficult. Note that these problems are central to this research. First, the spatially biased distribution of monitoring stations affects the creation of variogram clouds. In the case of Seoul, the 57 background stations are distributed across a 40km by 30km space but over half of them are distributed in Seoul. The areas that have sparse monitors may well have a higher uncertainty in fitting the results because there might be stations that are further away from each other but have similar values. Unexpected outliers (e.g. mountains) might exist between a pair of stations, and although they are close, this completely opposes the idea of spatial autocorrelation. Studies suggest that an appropriate number of sample points to generate a variogram model should be over 50 (Nyhan et al., 2016; Wong et al., 2004) or 100 (Li and Heap, 2014), but the spatial extent is also a critical factor to build a variogram.

Second, the user has to choose an appropriate theoretical semivariogram that best fits the empirical semivariogram. Theoretical models include spherical, circular, gaussian, exponential, stable, and many more. The fitting process can be automated but accuracy is not always guaranteed.

[1]The distance at which a spatial correlation exists.

[2]The upper limit of the semivariogram is the sill. A Sill minus a nugget is termed a partial-sill

[3]Represents short scale randomness or noise in the regionalised variable (Camana and Deutsch, 2020)

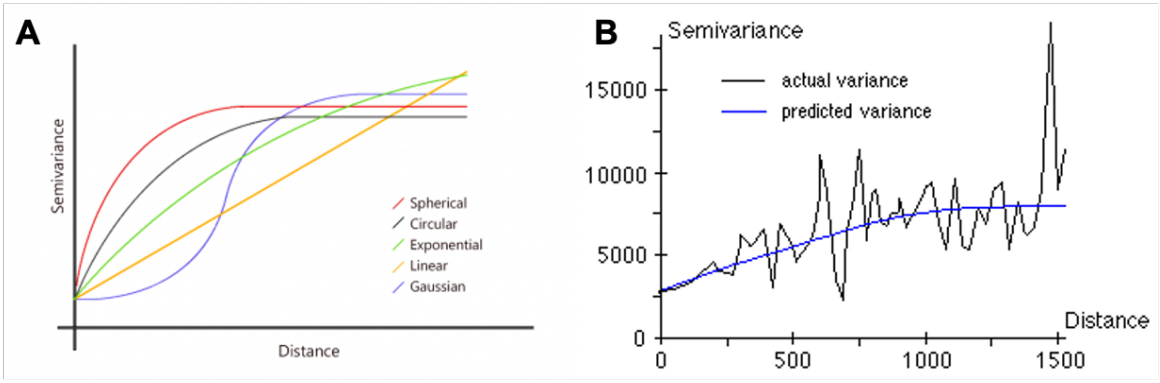


Fig. 3.7 (A) Types of semivariogram models (https://gisgeography.com/kriging-interpolation-prediction) and (B) the difficulty of fitting a theoretical model to an empirical model (http://resources.esri.com/help/9.3/arcgisdesktop/com/gp_toolref/spatial_analyst_tools/how_krige_and_variogram_work.htm)

Third, if the spatial extent is large, caution is needed when constraining the spatial lag (or cutoff). If the spatial lag is too short, then the model will not fully consider the variability of pollution concentration between distant stations. By contrast, if it is too long, then the model will consider stations that are less likely to have spatial autocorrelation. This study set the spatial lag of the variograms to 30km to consider all possible combinations between two points (see more spatial lag outputs in the Appendix).

The final point is the difficulty in controlling the nugget and finding the most adequate theoretical model. If the model does not have a nugget, this will yield measurement errors which generate a bull's eye map (see Figure 3.8). If the theoretical semivariogram does not fit into the points from each bin, then it will either have a bull's eye map or a simple, straight-contoured map. Bull's eye maps mean that the data is dominated too much by local station data, and it is supposed that the stations are not generally in a special location where the pollution levels are systematically higher than the surroundings. In other words, it is overfitted. The striped maps show that even closer distances have such a high variance that it cannot create a trend. This is also known as being under-fitted. The next section introduces examples of poorly and well-fitted semivariograms.

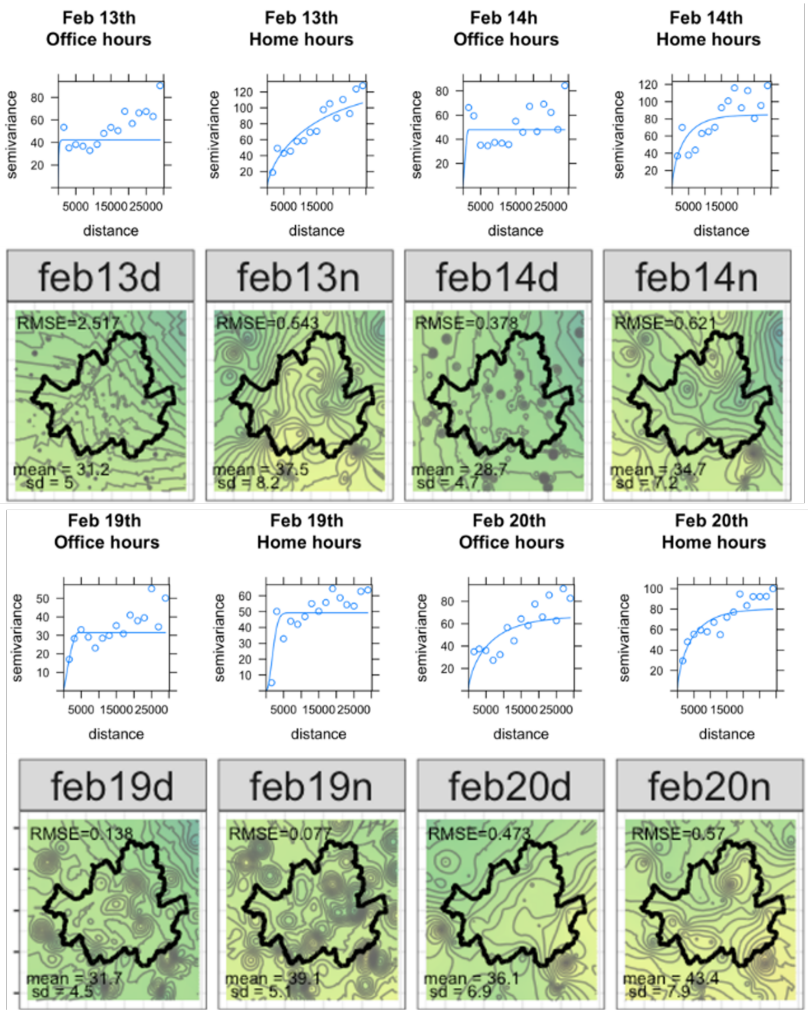


Fig. 3.8 An example of overfitted semivariograms that coerced the nugget to zero that led a bull's eye map (Office: day, Home: night, Units: ppb). Background NO_2 was analysed for this exercise.

Fitting Empirical Estimations to a Theoretical Semivariogram

This section attempts to fit a theoretical semivariogram to the empirical variogram cloud. To give an example from this chapter, a poorly fitted semivariogram (Feb 13th, Day) and well-fitted semivariogram (Feb 20th, Night) are introduced. Table 3.3 presented the semivariogram of NO_2 where the first three columns describe bin number, number of points (np), distance in metres, and the semivariances (gamma). In the first bin, 23 combinations between station points within 2.226km result in a gamma value of 52.12, and the second bin contains 80 combinations between station points within

Table 3.3 Semivariogram outcomes of NO_2 on 13th February 2014 (Left) and 20th February 2014 (Right)

Bin	np	dist(m)	$\gamma(Feb.13th)$	$\gamma(Feb.20th)$
1	23	2226.287	52.12715	56.05426
2	80	4453.986	24.66249	62.52822
3	120	7728.496	29.30227	61.02684
4	160	10491.6	28.80280	62.58777
5	170	13477.52	36.90700	66.68653
6	181	16570.44	33.20569	66.4433
7	191	19399.25	46.06388	82.67481
8	190	22336.83	46.35860	84.84201
9	132	25379.07	46.12718	95.03517
10	105	28408.96	43.90225	83.0663

Table 3.4 Semivariogram outcomes of NO_2 on 16th February 2014 (left) and 20th February 2014 (right)

Feb13 Day	psill	range(m)	kappa
Nug	24.03627	0	0
Ste	1449.7105	19,164.9	50
Feb20 Night	**psill**	**range**	**kappa**
Nug	56.53109	0	0
Ste	50.10143	27,771.23	5

4.453km that results in a gamma of 56.05. Compared to the 20th, the gamma value on the 13th dips on the second bin and then slowly increases, meaning that small scale noise dominates the closest bin. On the other hand, the semivariogram on the 20th had a nice fit.

Table 3.4 and Figure 3.9 illustrate the actual obstacles to finding a theoretical semivariogram. Both variograms have a chosen mathematical model `Ste` known as a Stein method, nugget 24.03 and 56.53, partial sill 1449.7105 and 50.10143, range 191,649m and 27,771.23m, and Kappa 50 and 5 respectively.

Since the predicted model depends substantially on semivariograms, kriging outcomes should therefore be rigorously validated after modelling, despite the time-consuming effort. However, during the summer, there were many days where the theoretical semivariograms did not fit into the empirical estimation. Referring to Figure 3.8, it was seen that the original data on the days of poorly fitted semivariograms

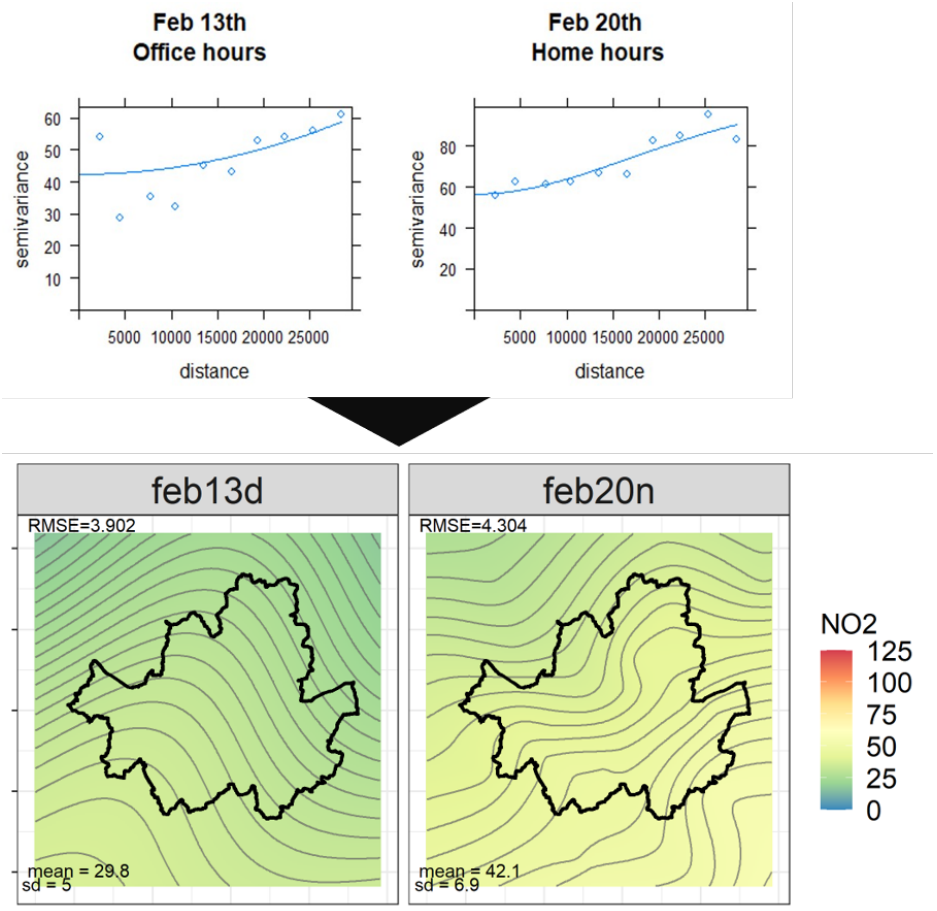


Fig. 3.9 Semivariogram examples (top) and interpolated outputs (bottom) of NO_2 on the day hours of the 13th and the night hours of the 20th in February. The theoretical semivariogram matches well with the empirical semivariogram on the 20th but did not fit well on the 13th. It can be speculated that there were high NO_2 variations between close stations. This is an evidence of small-scale variation that can be a weakness of spatial interpolation

were skewed closer to zero. The histograms itself proved that the normality assumption was already violated and thus had difficulties in finding a theoretical model because the NO_2 levels were very low during summer and day hours and had unexpectedly high NO_2 concentrations at some stations. To reduce the influence of the outliers, the fit was made with the logarithm of the data.

Figure 3.10 is an output of semivariograms between February 11th and 20th. Each graph adopted its own mathematical model that best accounts for the residuals, but the majority was using a Stein model, which best reacts to the abrupt changes between

the nearest points (Pebesma et al., 2019). The models that fit well with the residuals, such as the 11th, 13th, and the 17th home hours of February, show that the spatial autocorrelation was observed in those days and is reflected in a fairly smoothed output. On the contrary, models that either failed to create a good fit, such as 15th of February, resulted in straightened contour maps that missed local variations. All of the results are shown in the Appendix.

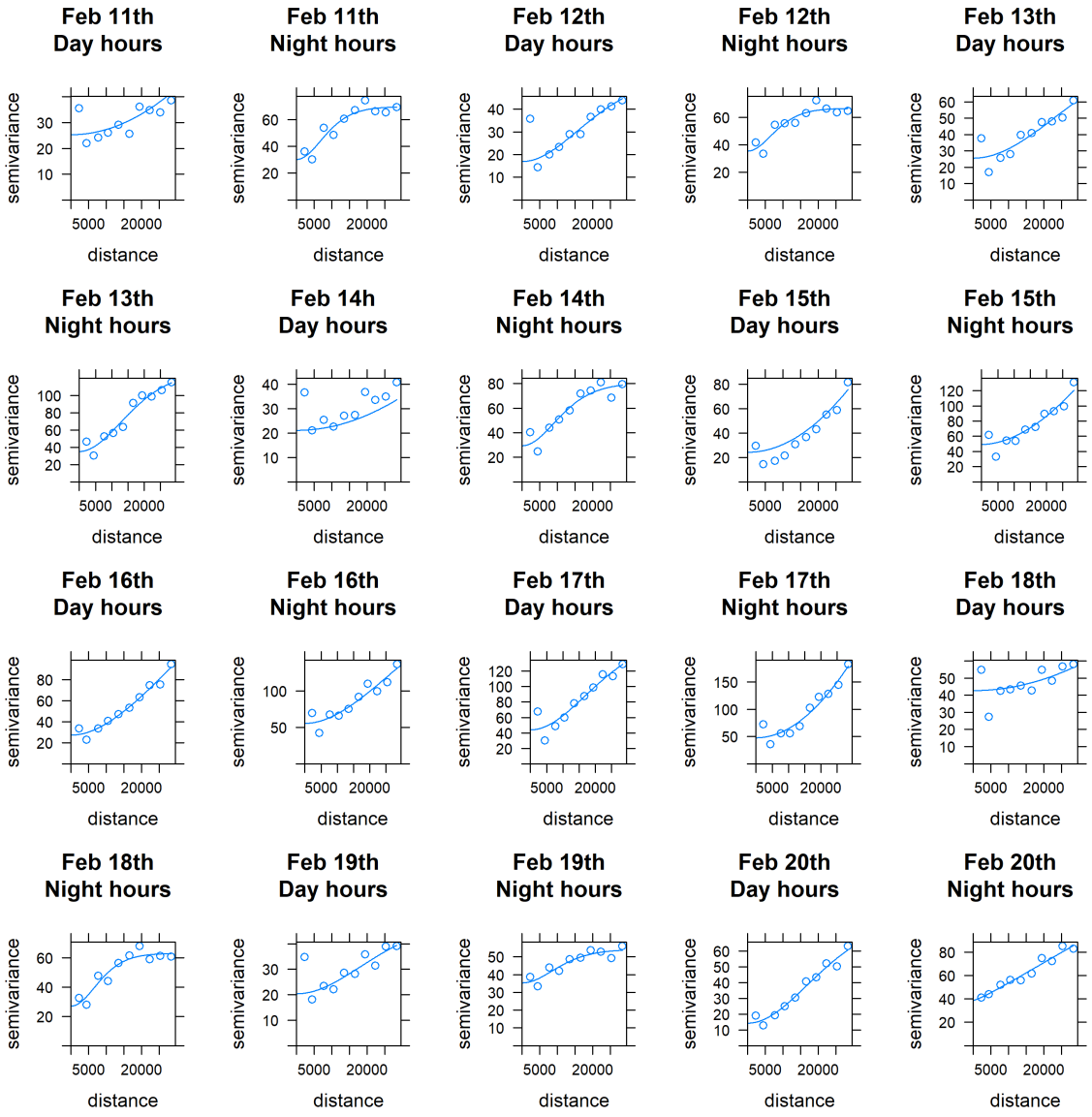


Fig. 3.10 Semivariogram of NO_2 in mid-February 2014. Because the mean distance between stations are around 29km (the farthest: 44km), this study confined the semivariogram's distance (spatial lag) within 30km

Ordinary Kriging: the Basics

Kriging is widely used in research because of the theoretical foundation that the weights decline as the distance between the data points increases (First Law of Geography), and the firm tool (i.e. semivariogram) that determine the coefficient of the equation (i.e. the number of neighbouring points, cutoff distance). Models such as the Inverse Distance Weight (IDW) method do not have a theoretical background nor a mathematical tool, but rather generate an estimation from a heuristic measurement. Thus, kriging might be more adequate for researchers who aim to estimate a continuous spatial extent with a few points and prefer a model with a theoretical framework.

Ordinary Kriging (OK) uses the acronym BLUE to describe its characteristics (best linear unbiased estimator). Kiš (2016) describes that, "Best is the word to minimise the variance of errors, linear because the estimates the linearly weighted combinations of data available, unbiased because it tends to have a mean residual at 0."

Conditions: Check Your Data

The most important task before the interpolation is to rigorously check the data structure. In theory, OK requires three assumptions, which are:

- Normality: the data needs to follow a normal distribution
- Stationarity: the data needs to be spatially stationary (see Figure 3.11)
- Trends: the data does not have any trends for OK[4], but does for UK (see Figure 3.12)

Most of the data used for this study tended to follow the normality assumption, although it varied by days and seasons. According to the principle, the data is aimed to be unbiased, which means it follows a Gaussian curve.

Stationarity means that the variance is almost constant in different locations on the map. For instance, if two random points are 30 metres apart, it is likely to have similar differences in the predicted values wherever they occur. However, the variance of atmospheric pollution may not always comply with the stationarity assumption.

[4]Systematic change in data across an entire study area, also known as drifts. e.g. North-South, East-West

This is because the wind eddies, built environments, or even the sampling filters inside the monitor enable a large variability (see Figure 3.10). Using non-stationary data can cause less accurate results for OK since spatial autocorrelation cannot be formed correctly in the unsampled sites, and the spatial error models can be wrong (Lark, 2009).

If the data follows the normality and stationarity assumptions and also does not have trends, then there is no problem to adopt OK. By contrast, if the data has trends, Universal Kriging should be considered.

Equations

In Equation 3.2, the principle of kriging is to predict a value $Z(s_0)$ at location s_0 using the observations of $Z(s_i) = Z(Z(x_1), Z(x_2), Z(x_3))$ and the calculation of the weighted sum of the data (Kumar, 2007).

$$Z(s_0) = \sum_{i=1}^{N} \lambda_i Z(s_i) \tag{3.2}$$

- $Z(s_i)$ = the measured value at the ith location
- λ_i = an unknown weight for the measured value at the ith location
- s = the prediction location
- N = the number of measured values

In OK, mean μ is assumed to be constant over the study space. The weight (λ) depends on: the 1) distance between measured points, 2) predicted location, 3) and the spatial weights utilised by the variogram model (Paramasivam and Venkatramanan, 2019). Based on the fact that OK has one constant mean, it will have a large variance depending on the distance from the data points. However, to ensure that the estimators are unbiased and the estimated values have minimal variation, the first condition is to check the mean residual to be 0,

$$E\{Z_v - Z_0\} = 0 \tag{3.3}$$

the sum of weights (λ) to be 1,

$$Z(s_0) = \sum_{i=1}^{n} \lambda_i = 1 \tag{3.4}$$

and use an external parameter (Lagrange multiplier) to minimise the variance of errors between the global mean and the error value. The Lagrange factor smooths the whole map by adjusting the sum of squares equal to 1.

$$\begin{pmatrix} \gamma(Z_1 - Z_1) & \gamma(Z_1 - Z_2) & \dots & \gamma(Z_1 - Z_n) & 1 \\ \gamma(Z_2 - Z_1) & \gamma(Z_1 - Z_2) & \dots & \gamma(Z_1 - Z_n) & 1 \\ \gamma(Z_3 - Z_1) & \gamma(Z_1 - Z_2) & \dots & \gamma(Z_1 - Z_n) & 1 \\ 1 & 1 & \dots & 1 & 0 \end{pmatrix} \times \begin{pmatrix} \lambda 1 \\ \lambda 2 \\ \lambda n \\ \mu \end{pmatrix} = \begin{pmatrix} \gamma(s_1 - s) \\ \gamma(s_2 - s) \\ \gamma(s_n - s) \\ 1 \end{pmatrix} \tag{3.5}$$

- γ: variogram values
- $s_1 \dots s_n$: real value at location 1 to n
- s: location where new value is estimated
- λ: Lagrange factor

If it was supposed that the data points were not randomly distributed and had trends, OK will ignore the trends and give the mean value, whereas in Universal Kriging (UK), the mean constant has a functional dependence to represent its local variation (Kumar, 2007). The local variation is also known as a *local trend* or *drift*. Each location s under universal kriging is recalculated considering the neighbouring values, which can be similar to a moving average.

Which Kriging method should I use? OK versus UK

To test the stationarity of the dataset, the author used Thiessen polygon maps (see Figure 3.11). Thiessen polygon maps are created by a set of sampled points by drawing lines through the mid-points of the lines joining nearest neighbours and keeping the line segments that form the smallest resulting polygon. Any location inside the polygon is closer to that point than any other sampled points (Law and Collins, 2019). The colour codes for each polygon area can vary by the attributes and the classification method. If the colour groups are presented in a random distribution (Figure 3.11, left – an example of a random distribution dataset), the data are likely to be stationary. Conversely, if the colours of standard deviation are clustered (Figure 3.11, right, modelled from this

study), then the data are non-stationary. Having tested many samples from this study, the outcomes showed small and large amounts of clusters on the Thiessen polygon maps. The result also has a drift (spatial trend). Polygons with light blue are centred around the core of the map whereas the comparatively more polluted polygons are observed at the edge of the map; this fits the justification of using UK. Details are described in the next section.

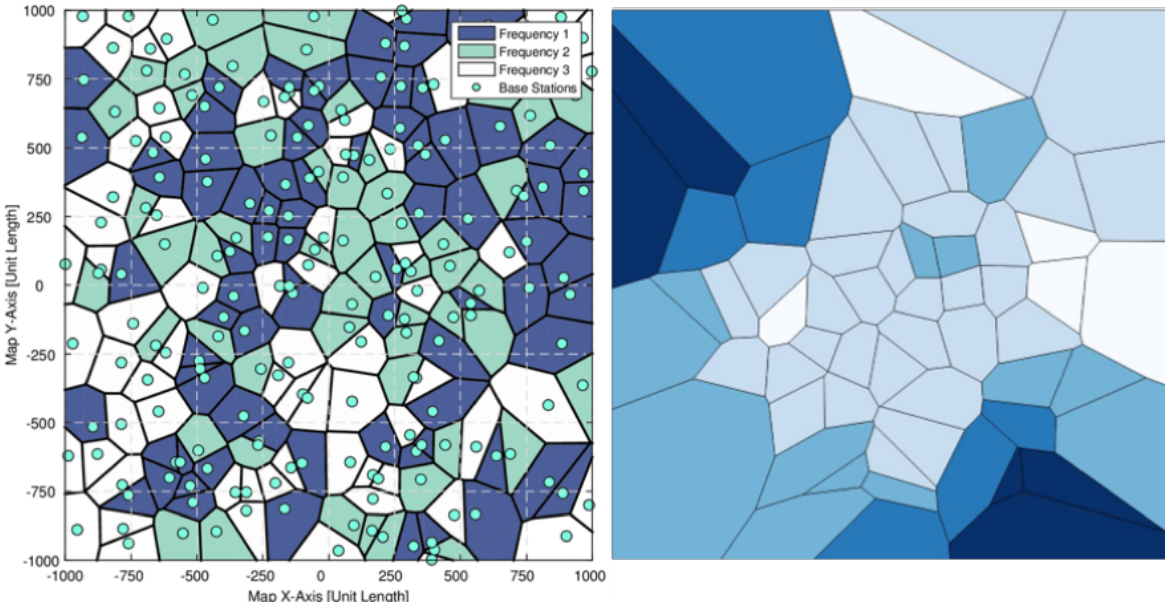


Fig. 3.11 Thiessen polygon maps that show the randomness between neighbours. The figure on the left is an example of a Thiessen polygon map where the values are randomly distributed colours (http://tiny.cc/voronoi). The right figure portrays a clustered distribution of Seoul's NO_2 measurement on 11th of February 2014.

Predicting a Kriged Surface with Universal Kriging

As mentioned above, it is difficult to find a parameter that meets the hypotheses of data normality, stationarity, and no drift effect because of the systematic change in the parameter value (Bárdossy, 1997). For instance, NO_2 is a parameter that is affected by traffic that shows a systematic change by the hours of the day, and PM_{10} also varies systematically by seasons; thus, it would be risky to be examined by OK.

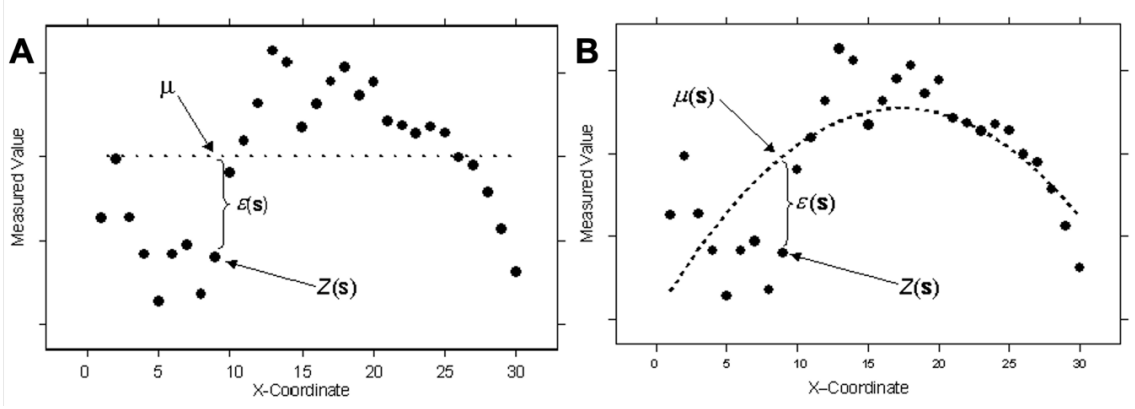


Fig. 3.12 An illustrative example of how (A) Ordinary Kriging and (B) Universal Kriging capture the mean value. The figure were copied from https://desktop.arcgis.com/en/arcmap/latest/extensions/geostatistical-analyst/understanding-universal-kriging.htm

A possible way to interpolate the data considering the drift effect is using UK. Figure 3.12 is an example of how OK and UK calculate the mean and the error values. The difference to OK is that it has a trend surface or drift (i.e. a global mean), and residuals (i.e. a global mean-local mean within the search boundary) that depend upon spatial coordinates (Fritz et al., 2009).

$$Z(s_0) = \sum_{i=1}^{N} \lambda_i Z(s_i) \tag{3.6}$$

- $Z(s_i)$ = the measured value at the `i`th location
- λ_i = an unknown weight for the measured value at the `i`th location
- s = the prediction location
- N = the number of measured values

Burrough et al. (2015) note that if x has a position in two dimensions in space, the spatial variation of any variable `Z` can be expressed as follows:

$$Z(x) = m(x) + \epsilon(x) \tag{3.7}$$

- m(x): a mean value which has a constant trend in UK
- ϵ(x): a stochastic component that can vary according to the region, i.e. a regional variable

Subsequently, the m(x) is composed of another equation:

$$m(x) = a_0 + a_1x_1 + a_2y_1 + a_3x^2 + a_4y^2 + a_5xy \tag{3.8}$$

- x_i: x coordinates of the `i`th point
- y_i: y coordinates of the `i`th point
- a_n: drift coefficients. If the degree is 1 (i.e. the trend is flat), then a_3=a_4=a_5=0. If the degree is 2 (see Figure 3.12, right), then the local trend is quadratic
- This equation is recomputed for every pixel

To obtain a minimum variance of error between the known and the unknown points ϵ(x), the Lagrange multiplier is applied in UK:

$$\begin{cases} \sum_{i=1}^{N} \lambda_i \gamma(s_i, s_j) + \sum_{k=1}^{P} \rho_k f_k(s_j) = \gamma(s_i, s_0)\ i = 1, 2, 3, \ldots,\ N \\ \sum_{i=1}^{N} \lambda_i f(s_i) = f(s_0)\ i = 1, 2, 3, \ldots,\ k \end{cases} \tag{3.9}$$

- $\gamma(s_i, s_j)$: semivariogram between the two points s_i and s_j
- ρ_k: Lagrange multiplier

Thus, the semivariance $\gamma(s_i, s_j)$ and the weight factor (λ_i) generate the variance of error shown below.

$$\epsilon_k{}^2 = \sum_{i=1}^{N} \lambda_i \gamma(s_i, s_j) + \sum_{k=1}^{P} \rho_k f_k(s_j) \tag{3.10}$$

- $\epsilon_k{}^2$: variance at the estimated point
- $\gamma(s_i,s_j)$: semivariogram values between the points

- ρ_{k}= Lagrange parameter (regional smoother) associated with the unbiased location

More arithmetic information is introduced in Malvić and Balić (2009) and Chung et al. (2019).

3.3.4 The Generalised Additive Model(Statistical Inference Model)

GAM: The Concept

The Generalised Linear Model (GLM) and the Generalised Additive Model (GAM) are regression methods that predict a response variable `Y` based on explanatory variables `X`. Compared to GLM, GAM uses combinations of linear and non-linear smoothings of explanatory variables to predict response variables. In other words, GAM is a semi-parametric model[5] as the assumptions are generous, and the relationships between response and predictors depend linearly on unknown smooth functions, thus can find their best fit (Wood, 2017). This flexibility can be compared to OLS regression or GLMs that are underpinned by strict parametric assumptions, i.e. average change in response variable is proportionate to the change in the parameters. Thus, GAM is very useful when the relationship between the dependent and predictive variables are expected to be complex, and not easily fitted to a linear estimate. Moreover, categorical predictors are also able to be included in the model, e.g. landcover types, gender. Here is the conceptual equation of GAM:

$$g(\mathbb{E}(Y)) = S_0 + s_1(x_1) + ... + s_{\mathrm{p}}(x_{\mathrm{p}}) \tag{3.11}$$

, where `Y` is the dependent variable of the index, $\mathbb{E}(Y)$ denotes the expected value, and `g()` denotes the link function that links the expected value and the predictor variables `x1, ... xp`. The terms of $\mathrm{s}_0 + \mathrm{s}_1(\mathrm{x}_1) + ... + \mathrm{s}_{\mathrm{p}}(\mathrm{x}_{\mathrm{p}})$ denote smooth and non-parametric functions. Each predictor has a relationship with the dependent variable that can be indicated as a scatterplot. Once the dots are split into sections, the linkage between sections, called *knots*, is matched with polynomial functions such as a piecewise linear regression, or smooth functions including loess, cyclic cubic, P-spline,

[5] A semi-parametric model is a statistical model that has parametric and non-parametric components. https://en.wikipedia.org/wiki/Semiparametric$_m odel$

or tensor product (Wood, 2017). Adding functions to knots to predict the link function is what we call an *additive* model.

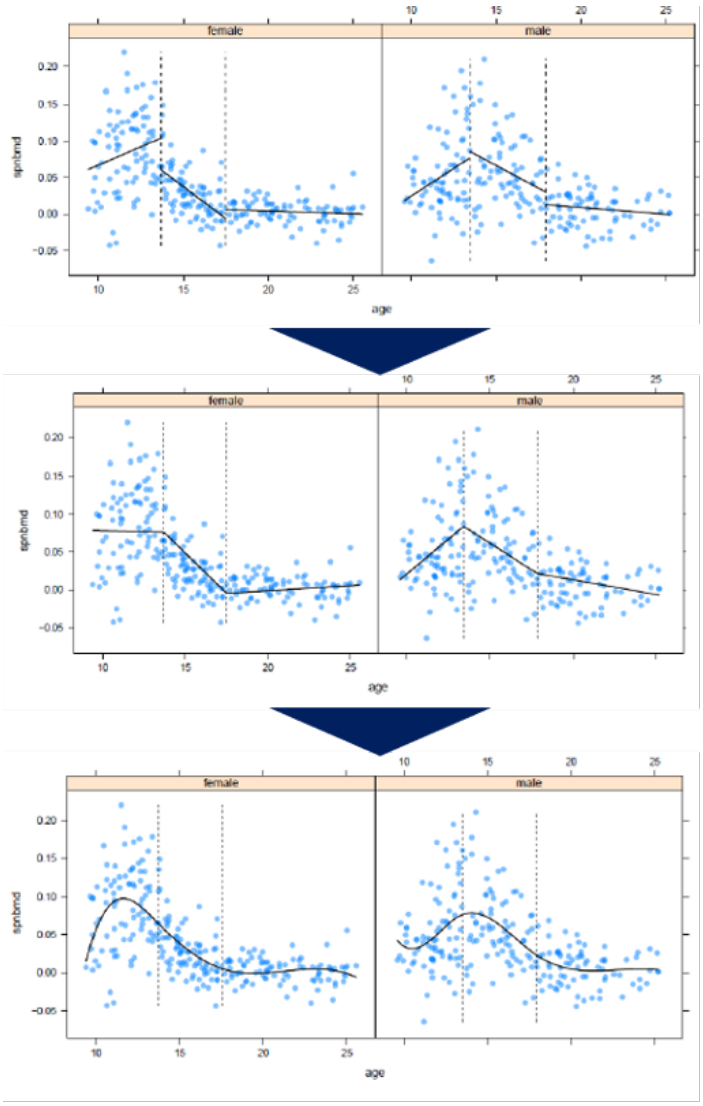


Fig. 3.13 An example of steps to get a smooth spline in GAM (adapted from Wood (2017)). To interpolate the points with a cubic spline, the first row finds the linear trend of the data points for each block. Here each block is distinguished according to the data distribution. Second row shows how unconnected lines are connected. The connected points are termed as knots. Since the lines within each block does not fully reflect the data points, the model finalises the equation by applying the cubic function. Note that choosing high degree of smoothing might result in overfitting the outcome. For theoretical information, please refer to Wood (2017), and for explanations with coding and visualisations please refer to https://m-clark.github.io/generalized-additive-models/

Choosing a Smoothing Function for Covariates

In the `mgcv` package, there are various smooth functions that can be used to represent single or multiple penalties to control the degree of smoothness (Wood, 2017). Thin plate regression splines are a type of smoothing spline used for visualising the complex association between predictors and responses on a continuous surface[6]. The advantage of using thin-plate splines is that GAM does not require any *a priori* knowledge to the form of data, and because of the multidimensional output, two continuous predictors can be used to estimate a single outcome.

However, a downside is that it is difficult to visualise the error rates. The R package takes this as a default spline with codes written as `bs='tp'`. Other smooth terms include Duchon splines, cubic regression splines, P-splines, random effects, Markov random fields, and so on. These have their own assumptions and required data structure. This study uses the default spline.

Choosing the 'Wiggliness' (Complexity) of the Smoothing

This study used the Restricted/Residual Maximum Likelihood (REML) as a smoothing method. However, before REML is elucidated, there needs to be some background knowledge of Maximum likelihood (ML) and how ML is used in GAM. ML is a method that maximises the likelihood of observing the weights measured. Instead of the normal distribution that has the mean in the centre and has an equal tail, ML finds the curve that best accounts for the distribution of the data. As ML requires both fixed and random effects, many points are likely to get a better prediction.

Unlike ML, REML does not consider the fixed effects but rather controls the estimate of variance parameters. When the REML estimates are used, the fixed effects – a linear combination of unknown parameters – are taken away, thus the residuals will have a mean of 0. This is slightly different from the ordinary ML where both parameters of the fixed effects and the residuals are taken into account. Thus, when there are a few points used in the model the trends are easier to draw with REML, but REML and ML become negligible when the sample size is large (Korner-Nievergelt et al., 2015).

[6] http://tiny.cc/thinplatespline

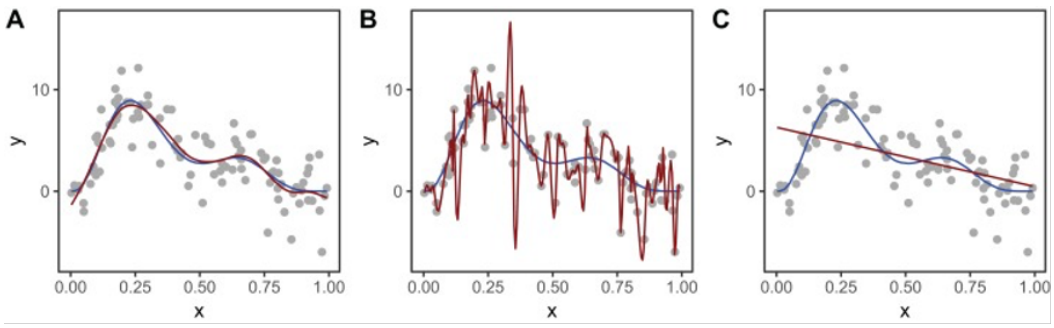


Fig. 3.14 An example of choosing a smoothing parameter: (A) REML, (B) zero smoothing parameter, (C) large smoothing parameter that ends up removing the wiggliness (Pedersen et al., 2019)

Choosing a distribution curve: Maximum likelihood (ML) attempts to find the optimal way to fit distribution for the data. The curve can be in a normally distributed curve, left-skewed, or right-skewed, exponential, gamma, or possibly something else. Choosing an accurate distribution curve is a means to better predict the phenomenon in a certain interval. In Figure 3.14, for example, the gamma curve will fit the distribution.

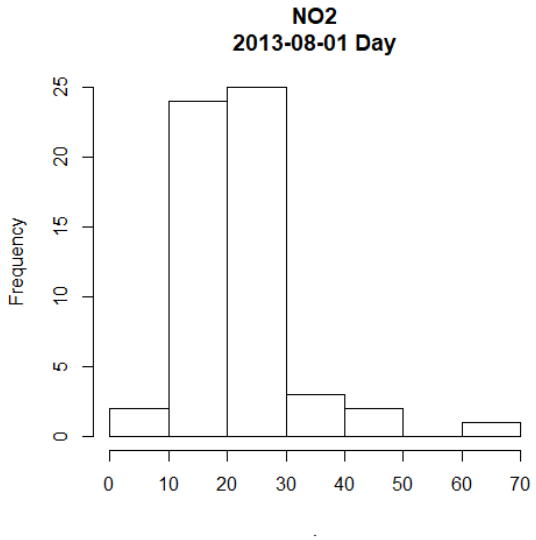


Fig. 3.15 NO_2 distribution on the 1st of August 2013 in Seoul background stations

One way to check the model fitness is to use residual plots. Using the data from Figure 3.15, the residuals are distributed in two types: ML (left) and REML (right) (see Figure 3.16). In Figure 3.16A, the GAM model with Gaussian distribution showed

that the residuals were biased, and therefore might underpredict the values for spatial interpolation. The residual plots on the top-right and bottom-left show that the residuals are not evenly distributed and skewed. An alternative was to use a Gamma curve that considers the left-skewed data distribution (see Figure 3.16B). Although small the QQplot shows that the distribution is slightly closer to the one-to-one line, it is not skewed too much in the top-right nor bottom-left plots, and is slightly underestimating the response values.

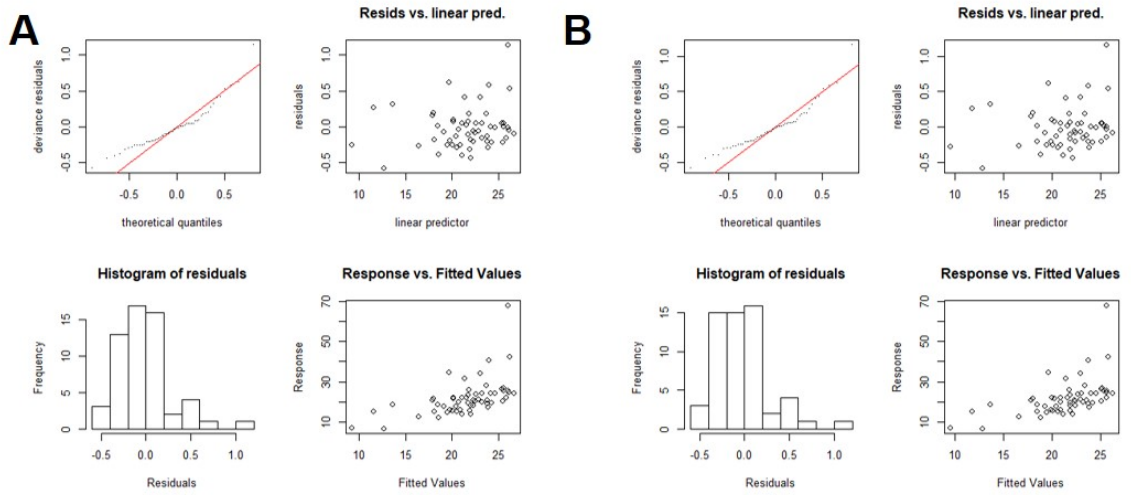


Fig. 3.16 Four plots checking residual distribution with (A) Gaussian and (B) Gamma methods. For each method, the residuals are illustrated in a QQplot (Quantile-Quantile plot), residual-linear prediction plot, histogram, and response-fitted values. The upper left plot is the QQplot which checks the normality of the data. While the Gaussian method has an outlier at the top-right, the Gamma method residual has a marginally better fit. The upper right plot is another residual plot to consider how residuals are well distributed as the mean indicator (predictor) increases. Evenly distributed residuals are considered a better model fit. The bottom-left histogram shows residuals by frequency. The best fit has a normally distributed curve. The lower right panel shows a trend of how well the fitted values match the response value: one-to-one lines mean a perfect fit.

Including knots to adjust the wiggliness

Each predictor can manipulate its own wiggliness by including or excluding knots. Including more knots can lead to overfitting, while less knots can over or underestimate values. the GAM is penalized if it uses too many degrees of freedom (Wood, 2019).

Wood (2019) also states that:

"The trade-off in selecting a smoothing parameter is whether to opt for less bias but with much variance, or smoother curves (k < 5) with higher bias but less variance. Mathematically speaking, this previous sentence can be translated as minimising residual deviance (goodness of fit) while maximizing parsimony (lowest possible degrees of freedom)."

Since the deviance of the smooth functions can fit, a modeller has to be conscious of controlling the wiggliness. Noam Ross, an online instructor of *Nonlinear Modeling in R with GAMs*[7], also emphasised that modellers should take into account two things when fitting a nonlinear model: closeness to data (avoiding underfitting), and not fitting the noise (avoid overfitting). Possible alternatives are to check whether the knots are too many or few by using `gam.check`[8] or `qq.gam`[9] functions in R.

Testing the Model's Robustness

GAM supports a number of indices to test its robustness. `R-squared` is the coefficient of determination scaled between zero and one, `edf` is the estimated degrees of freedom (the larger the number, the more wiggly the fitted model). `Ref.edf` is reference degrees of freedom[10], F is the probability that the null hypothesis for the model is true, and the p-value is the statistical value against the null hypothesis. Wood (2017) noted that the p-value has to be used carefully interpreted because it is normally reliable when the model is unpenalised. The `p-value` here is based upon the `edf` and the rank of the covariate matrix for the coefficients on a particular smooth, which can vary by knots and model iteration.

In terms of the model's goodness-of-fit, there are many indices such as `AIC`[11], `-REML`[12], or `RMSE`. The `mgcv` package contains `AIC` and negative `REML` scores by default.

[7] https://noamross.github.io/gams-in-r-course/

[8] It gives four types of residual plots and some diagnostic tests of whether the dimensions choices are adequate

[9] Residual qqplot of GAM

[10] These are a bit of a throwback, and are not very useful – they are reference degrees of freedom used in computing test statistics and the p-values, but since the null distributions are non-standard the reference DoF is not very interpretable.

[11] Akaike Information Criterion: an estimator of the relative quality of statistical models for a given set of data. AIC deals with both the risk of overfitting and the risk of underfitting. It is useful for the goodness-of-fit

[12] Restricted Maximum Likelihood: REML estimators maximize only the portion of the likelihood that does not depend on the fixed effects

If the `AIC` scores of GAMs are higher than GLM which means worse fit, it is more useful to use GLMs since GLMs are easy to interpret.

GAM with a Univariate Smoother

Figure 3.17 accounts for a nonlinear relationship between the `X` and NO_2 during the daytime hours of the 1st of August 2013. The family and link function consider Gaussian distribution which means that it assumes the data to follow normality. The formula shows that the dependent variable NO_2 will response from the `s(X)` or smoothed X variable with 10 knots. The effective degree of freedom (edf) with 1.99 suggests that the knots and smoothness have a quadratic curve as a nonlinear effect. The wiggliness of 1.99 fits the balance between an undersmoothed and oversmoothed fit (Wood, 2017). Although the p-value is not a strict cutoff for hypothesis testing, the value .254 suggests that there is little evidence that rejecting the null hypothesis of the smoothing term is necessary or that a simpler model would fit (Young et al., 2011). Another thing to note is that it accounts for only a little variance in NO_2 levels, with an adjusted R^2 of 4.37%.

```
Family: gaussian
Link function: identity
Formula:
no2_8_01_day ~ s(X, k = 10)
Parametric coefficients:
            Estimate Std. Error t value             Pr(>|t|)
(Intercept)   21.901      1.184   18.49 <0.0000000000000002 ***
---
Signif. codes:  0    ***    0.001    **    0.01    *    0.05    .
   0.1         1
Approximate significance of smooth terms:
      edf Ref.df     F p-value
s(X) 1.99  2.501 1.274   0.254
R-sq.(adj) =  0.0437   Deviance explained = 7.77%
Model rank = 30 / 30 -REML = 203.28  Scale est. = 79.932    n = 57
```

Listing 3.1 R code for GAM output

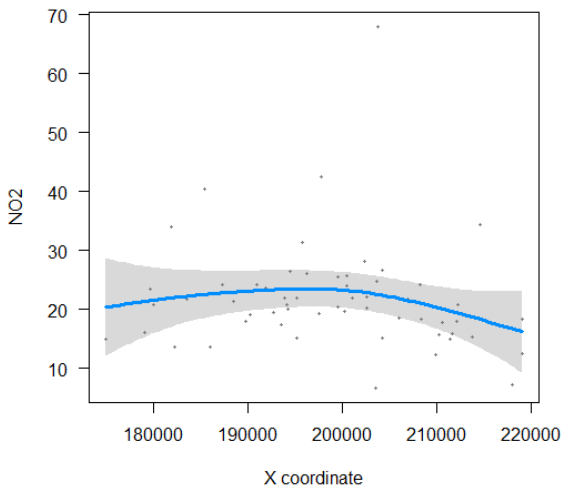


Fig. 3.17 A non-linear relationship between the X coordinate and NO_2 concentration

How Variable Interaction Works as Geospatial GAMs

To extend the smooth functions of GAM on a geospatial aspect, the interaction of multiple variables can possibly work. GAM will fit models of pollutants associated with X and Y, and visualise those surfaces in 2D or 3D. Here is an example outcome of NO_2 associated with X, Y coordinates on the day hours of August 1st 2013.

```
Family: Gamma
Link function: identity

Formula:
no2_8_01_day ~ s(X, Y, k = 15)

Parametric coefficients:
            Estimate Std. Error t value             Pr(>|t|)
(Intercept)   21.542      1.052   20.48 <0.0000000000000002 ***
---
Signif. codes:  0    ***    0.001    **    0.01    *    0.05    .
   0.1         1

Approximate significance of smooth terms:
         edf Ref.df     F p-value
s(X,Y) 6.028  8.077 2.998 0.00751 **
---
```

```
17 Signif. codes:  0    ***    0.001    **    0.01    *    0.05    .
       0.1          1
18
19 R-sq.(adj) =  0.137   Deviance explained = 33.6%
20 -ML =  191.7   Scale est. = 0.13525   n = 57
```

Listing 3.2 Interactions between variables

The form of the output looks similar to the univariate model, but the difference is shown in the coefficients, rank, R^2, and on the deviance. The first line of the code assumes that the predictors follow a gamma distribution and the link function of the smoother identity means that no transformation was made to fit the model.

The formula indicates that NO_2 is predicted by the smoothed X and Y with 15 knots. The effective degree of freedom (EDF) at 6.028 indicates that the wiggliness goes above six dimensions to fit the response variable (R^2 .13, Deviance 33.6%). Here, p-value of .00751 roughly informs that the null hypothesis of a zero effect of the indicated spline is rejected; in other words, wiggliness is therefore statistically acceptable for this model. Wood (2017) notes that the approximation increases when the EDF is 1.00 which is a linear (unpenalised) model. However, he also mentions that we should be careful when interpreting any coefficients by a formula because the p-value can easily change from an extra input of knots which can overstate the result. Further information about overfitting and underfitting will be explained in Section 3.3.5.

Figure 3.18 is a 2D map of a spatial GAM. With the interactive smooth terms in the formula, GAM considers x-axis and y-axis as longitude and latitude. Figure 3.18 shows the contours of predicted values and the standard errors on a 2D surface where the black lines are the predicted values, red dotted lines for -1 standard error, and green dotted lines for +1 standard error. Using the GAM approach, the interpolated contour maps of NO_2 on the 1st of August 2013 is shown in (C). The contours on the map get wigglier as the k increases and vice versa when it decreases. Figure 3.18A is entitled `s(X,Y,6.3)`, which means it used a smooth interactive function with an `EDF` of 6.3. This plot used `k'` = 30. The black contours show the main estimate and the dotted contours show standard errors (red: -1 se, green +1 se).

The mean NO_2 of this map is 20.094ppb, and the black contour in the centre draws the mean estimates. The other black contour on the top right indicates -10ppb from the mean. The green dotted contour in the centre is a +1 `se` from the mean, and the red dotted contour just outside the black contour illustrates -1 `se` from the main

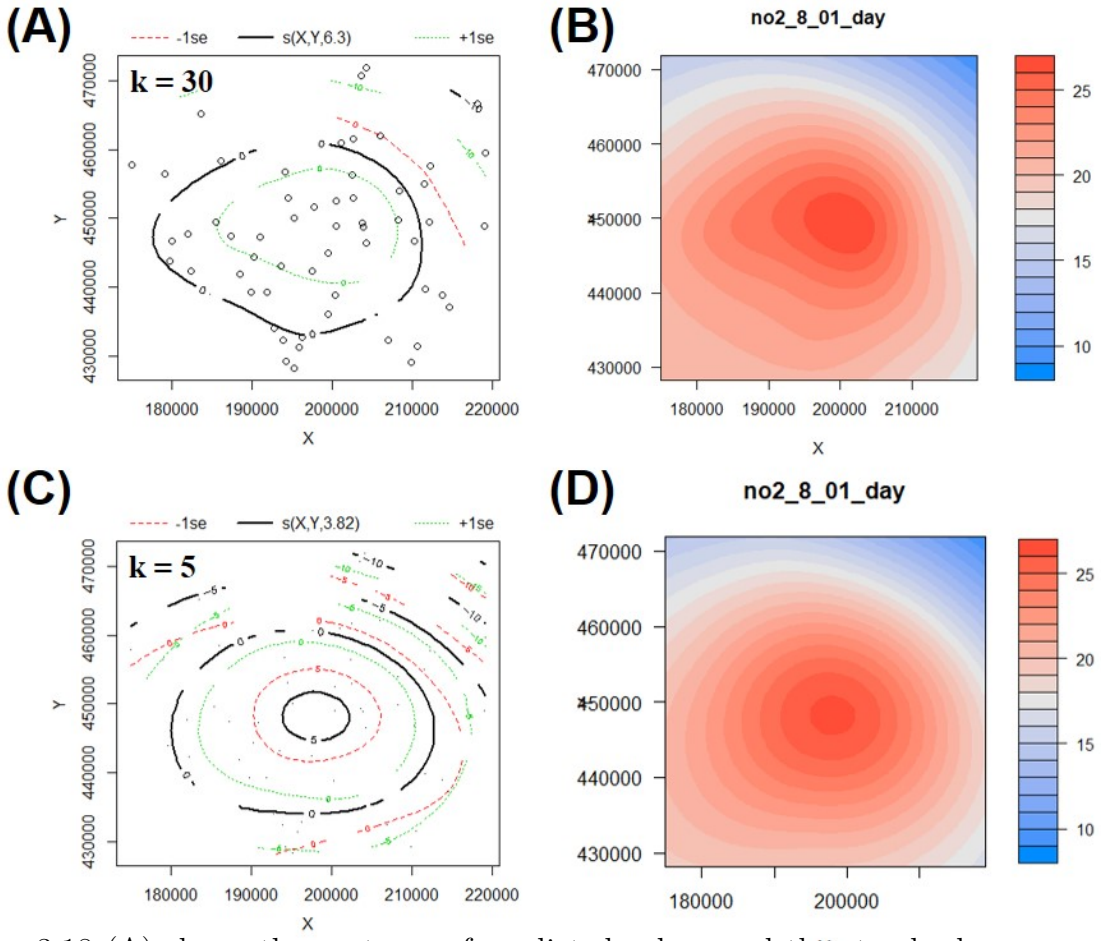


Fig. 3.18 (A) shows the contours of predicted values and the standard errors on a 2D surface where the black lines are the predicted values, red dotted lines are for -1 standard error, and green dotted lines are for +1 standard error. Using the GAM approach, the interpolated contour maps of NO_2 on the 1st of August 2013 is shown in (C)

estimate. The black contour in the top right corner has a -10 difference from the mean, with a green contour nearby. Figure 3.18B draws a contour plot out of Figure 3.18A, which has a high concentration in the centre at 26.2ppb and the lowest at 6.25ppb at the top right corner of the city.

Figure 3.18C has 4 knots with an `EDF` of 3.82. Compared to the higher `k'`s, the plot is slightly rounder with multiple ring buffers (Figure 3.18D). The `AIC` value of `k'`=40 was 384.8813, while `k'`=5 was 318.6873, meaning that the higher node model has a better fit. However, from Wood (2017) suggestion, a better model fit can be made somewhere between `k'`=5 and `k'`=10 given that the high number of `k'`s can lead to overfitting.

Why GAM's p-value is Different to OLS' p-value

The p-value for linear regression is a threshold to assess whether the null hypothesis should be rejected or not. p-value <.05 works when the conditions of normality of data distribution, independence between variables, and homoscedasticity (homogeneity of variance follow a linear trend. But what does the p-value mean here?

According to the explanation from Wood (2017), the smooth term attempts to find the best fit (from a Bayesian estimation) that can lower the p-values by iterating multiple times. However, Coupé (2018) mentions that the p-value of a smoothed variable should be interpreted cautiously and perhaps conclude as "slightly significant". From the last line of the direct quote below, Wood (2017) answers this cautiousness as not considering the uncertainty in the parameter estimates. In his book, he also mentioned that the p-values of the smoothers are sometimes substantially too low during the degree of penalisation, and this can be a matter of choosing the smoothing method, e.g. REML, ML, AIC, GCV. Consequently, the p-values are reliable for the model when the model is less penalised (if it follows an OLS trend).

3.3.5 Caveats to the Selected SI Methods

Exclusion of important parameters

Although both SI methods have the advantage of understanding the broad trend, the major drawback, particularly with pollution exposure, is that it does not consider various built environments (e.g. roads, bridges, buildings, parks) nor physical environments (e.g. mountains, weather) that cause discrepancies in our real world (Hoek et al., 2008). Moreover, Kriging outcomes that only consider pollution fields can underestimate the spatial variation of locations that are further away from the stations.

(An)isotropy

Spatial interpolations can bring different outcomes by directions. For example, areas with westerly winds can give different results. An anisotropic effect was revealed to be useful compared to an isotropic model in estimating surface-level wind speed at a (European) continental scale (Friedland et al., 2017), however, it did not suggest a great difference on a national level for wind speed in the UK (Fowler and Ekström,

2009), nor in a precipitation study in Finland (Aalto et al., 2013). Despite not using this approach for this study, this section describes how anisotropic effects account for directional influences.

For anisotropic Kriging, one can decide the degree of anisotropy angles to perform a variogram. However, the downside is that the user has to define the angles to examine the effect in a particular direction. GAM uses a tensor product to penalise different smoothing bases for different directions. Since the model smooths splines in multi-directions, it requires a form of matrices to create outcomes, which looks like a 3D image.

However, this study decided not to use anisotropy modelling. After testing degrees in eight directions (45°, 90°, 135°, 180°, 225°, 270°, 315°, 360°) on sample dates, the author observed that the variogram cloud fitted on semivariogram models only on a few days. This is because pollutants are likely to vary by land use and by localised effects such as traffic.

(To avoid) overfitting and underfitting

Overfitting and underfitting issues come alongside spatial interpolation because the areal prediction is made from a certain number of points. While underfitted maps appear to be quite simple, the overfitted maps will have very complex contours. This study questions how to avoid over- or underfitting but improve accuracy, and will be answered differently by each method.

In Kriging, A semivariogram is the fitting model for the kriging prediction, and many models can fit into your data, e.g. spheric, gaussian, or exponential. Since these conceptual models have their own shape, the chosen model depends on the location of the points and how the data is distributed. Unlike previous studies (Wong et al., 2004; Zhang and Shen, 2015), where annually-averaged pollutants were used, this study finds semivariograms from 12-hour aggregated data which is quite unusual to get a good fit, as the monitoring stations are biasedly distributed – more concentrated in the city centre than the outskirts – and the pollution levels tend to vary over very short distances that make it harder to meet the assumptions of spatial autocorrelation.

If there was a kriged map that has bull's-eyes or spirals that are the locations of stations, then it can be presumed as overfitted. In the semivariogram, *the nugget might be lower than the closest point.* On the other hand, if the map had contours that have

multiple strips, which shows underfitting, then it can be presumed either to have a nugget that is higher than the first point of the semivariogram, or the variogram models are far from the points. Hence, the best way to avoid overfitting and underfitting in kriging is to adjust the nugget value (intercept) closer to the first point of the semivariogram, and also find a suitable model that fits the semivariogram points.

In GAM, When using penalised regression smoothers, it is certain that the number of knots k' needs to be chosen. Adding a k' will increase the wiggliness to the model, but the point here is that the exact selection of k' is generally not that critical (Wood, 2017). Wood also adds that the accuracy of k' is very unlikely to be worthwhile if the EDF is far from k-1', but a check is required if the EDF gets close to k-1'. Wood (2017) also mentioned that an increment of k' will proportionately allow an increase to EDF, but these minute change will not affect the difference to the outcome.

To avoid overfitting or underfitting, GAM checks whether the residuals follow the normality or how equally variable they are (see examples in Figure 3.16). An ideal figure will have a nice fit with the red line of a QQplot, a nice random spread on the Resid vs linear pred plot, a bell curve on the histogram, and a perfect corresponding between the response and fitted values. In the figures below, the residuals tended to have a left-skewed distribution, and a few outliers in the residual plot, which in turn underestimated a certain number of points in the final result.

3.3.6 A New Proposed Method: Adding Road Effect to Background Pollution Levels on Spatial Interpolation Outputs

This section introduces a new approach to an additional road effect that can be used together with the above spatial interpolation outcomes. Land use regression (LUR), which is a regression-based method that includes the spatial aspect of infrastructural, meteorological variables to predict (long-term) pollution outcomes. For example, roads within 10m and wind speed showed the highest fit for the annual NO_2 value. LUR has a long history in dealing with pollution exposure with simple regression approaches and visually intuitive outcomes (Hoek et al., 2008). However, the model's strict assumption of aggregating and transforming variables in a single static snapshot can only account for one temporal aspect and neglects seasonal effects. Recent research from Lee et al. (2017) and Naughton et al. (2018) employed daily and hourly NO_2 in the model,

however, they had to aggregate it to an annual mean to keep the temporal scale correct with the other variables, e.g. annual average daily traffic, land use, and real estate. This will neglect the impact of the daily impact of pollution sources, particularly near roads. Also, if LUR is modelled for more than two time series, then the variables that were significant in the previous period might not be important in the next period, which does not make any sense. For example, a 100m or 200m road buffer that was statistically significant in 2015 might not be significant in 2019.

Fluid dynamic models that contain complex chemistry and meteorological mechanisms can produce high-quality pollution maps at the street canyon scale (Beevers et al., 2013; Di Sabatino et al., 2008; Tonne et al., 2008). However, technically speaking, the expensive computational job only processes a small area at once, which requires much effort to mosaic the outputs into an integrated map – more than twenty simulations needed for the size of London and takes around a week to finish the simulation. Even then it is hard to match these models against the station measurements.

Hankey and Marshall (2015) used a mobile monitoring method to improve on-road pollution measurements. The study was conducted in the urban area of Minneapolis in Minnesota during two rush hours, 7-9am and 4-6pm for two months from August 14th to October 16th 2012, with the pollution sensor attached to a bike wagon. Although the model improved by assembling finer spatial resolution data as well as finer temporal scales that led to more dynamic outcomes, it gave poor outcomes when the traffic congestion was extremely high or low.

To overcome these limitations, an alternative approach is to 1) measure a 12-hour average of background stations as well as roadside stations, 2) compare the ratio of roadside to the background, and then 3) apply the ratio to road layouts. As mentioned above, roads are classified as general roads and highways regarding the variance of pollution dispersion by road types. As a brief example, Seoul's background measurement of NO_2 in June, July, and August in 2013 was 25.12ppb, while the roadside was 43.44ppb and the (urban) highway was 44.56ppb. The ratio between the background and the roadside was 1.78 and between the background and the highway was the same at 1.84. As this is a first attempt to use ratios on spatial interpolations, the half-day ratio will be applied universally to the roads and highways irrespective of traffic flows.

Applying road:background ratios to GIS road data

Ratios of background:roadside and background:highway were calculated temporally for NO_2 and PM_{10} (see Figure 3.19 and Figure 3.20). Within each figure, graphs A and C represent the averaged concentrations of 57 background stations, 19 roadside stations, and two highway stations for summer 2013 and winter 2014, while graphs B and D represent the ratios of road versus background and urban highway versus road for summer and winter.

For NO_2, the overall concentration of background, roadside, and urban highway stations in summer were 25.12ppb, 43.44ppb, and 44.56ppb, while the mean level in winter was 40.88ppb, 52.71ppb, and 53.58ppb respectively, which in percentiles were 62%, 21%, and 20% higher than summer. The fluctuation range varied tremendously between day and night due to the photochemical effects from sunlight that accelerated the decrease during the daytime, although the range was greater during summer possibly due to longer sunlight hours. Besides this, the distinction of NO_2 between seasons and station types was noticeable. In summer, NO_2 levels of both road and urban highway stations were consistently 5-20ppb were higher than background results, which can be interpreted as an effect of high traffic volume. There was also a difference between stations in winter albeit a small number.

With regard to ratio calculations between station types, a considerable difference was observed between day and night. In summer, ratios of averaged data from road stations and urban highway stations were 1.78 and 1.83 times higher than background stations; this can be compared to winter ratios at 1.33 and 1.35. As mentioned above, the difference might have been made by photochemical effects during sunlight hours, but more weight is added to traffic volume. This can be proven by the large contrasts of an urban highway-background ratio where traffic is heavier during daytime and lightens after peak hours. The concentrations can increase during winter, but as the background level increases simultaneously the ratio is relatively lower.

Compared to NO_2, PM_{10} had a larger range in the annual spectrum rather than that of the daily range. It was particularly observable in December where the levels breached the annual average of 100µg/m^3. There were smaller differences seen by station types over the whole period, except in February 2014 where the contrast was prominent. As with NO_2, the effect of roadside pollution will not be reflected significantly in the overall map in low concentration days.

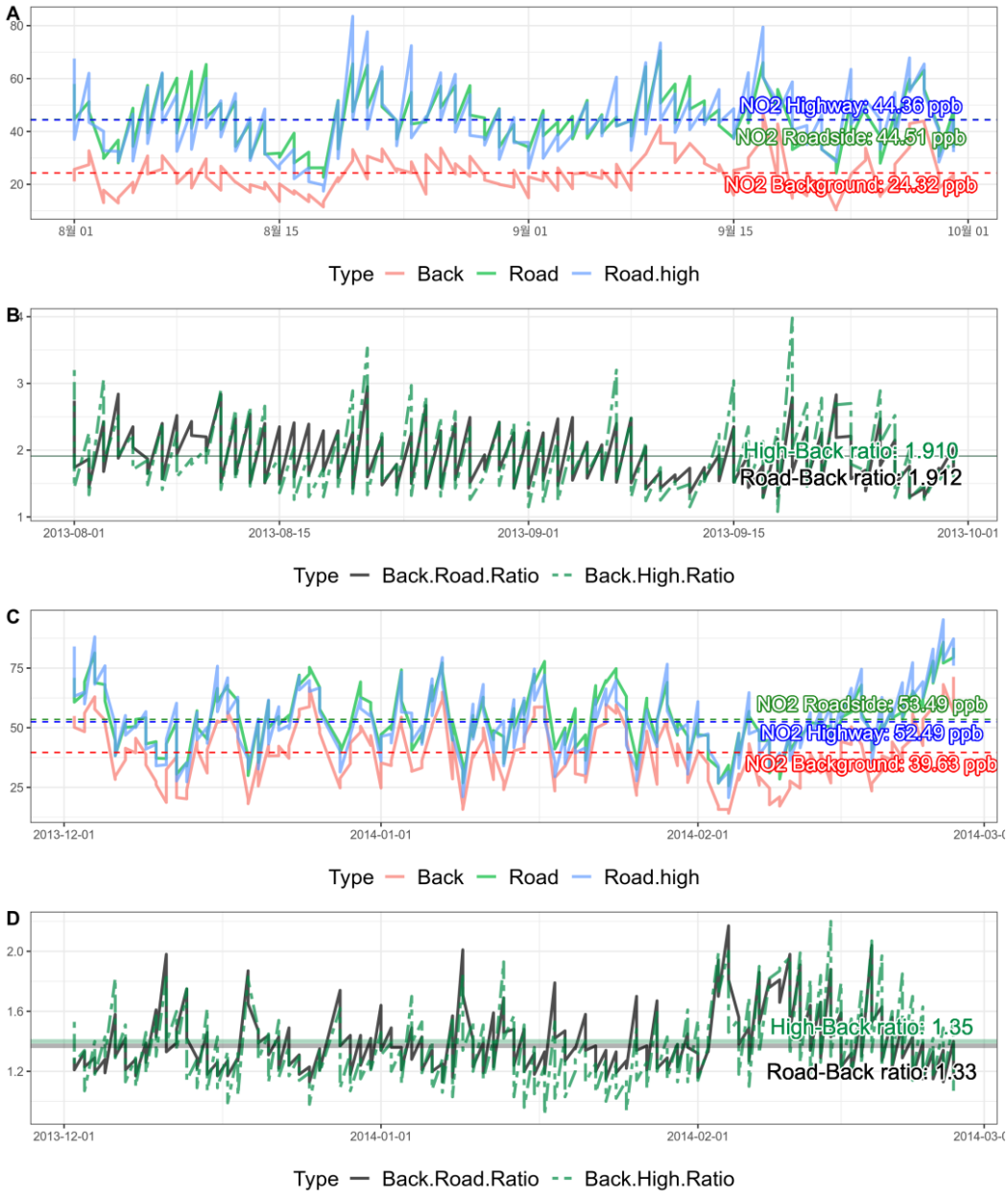


Fig. 3.19 NO_2 ratio measured by every 12-hour average in late July-September 2013 and December-February 2013-2014. Graphs A and C are pollution levels of NO_2, and Graphs B and D are ratios between aggregated roadside and background stations, and between aggregated urban highway and background stations by a half-day average

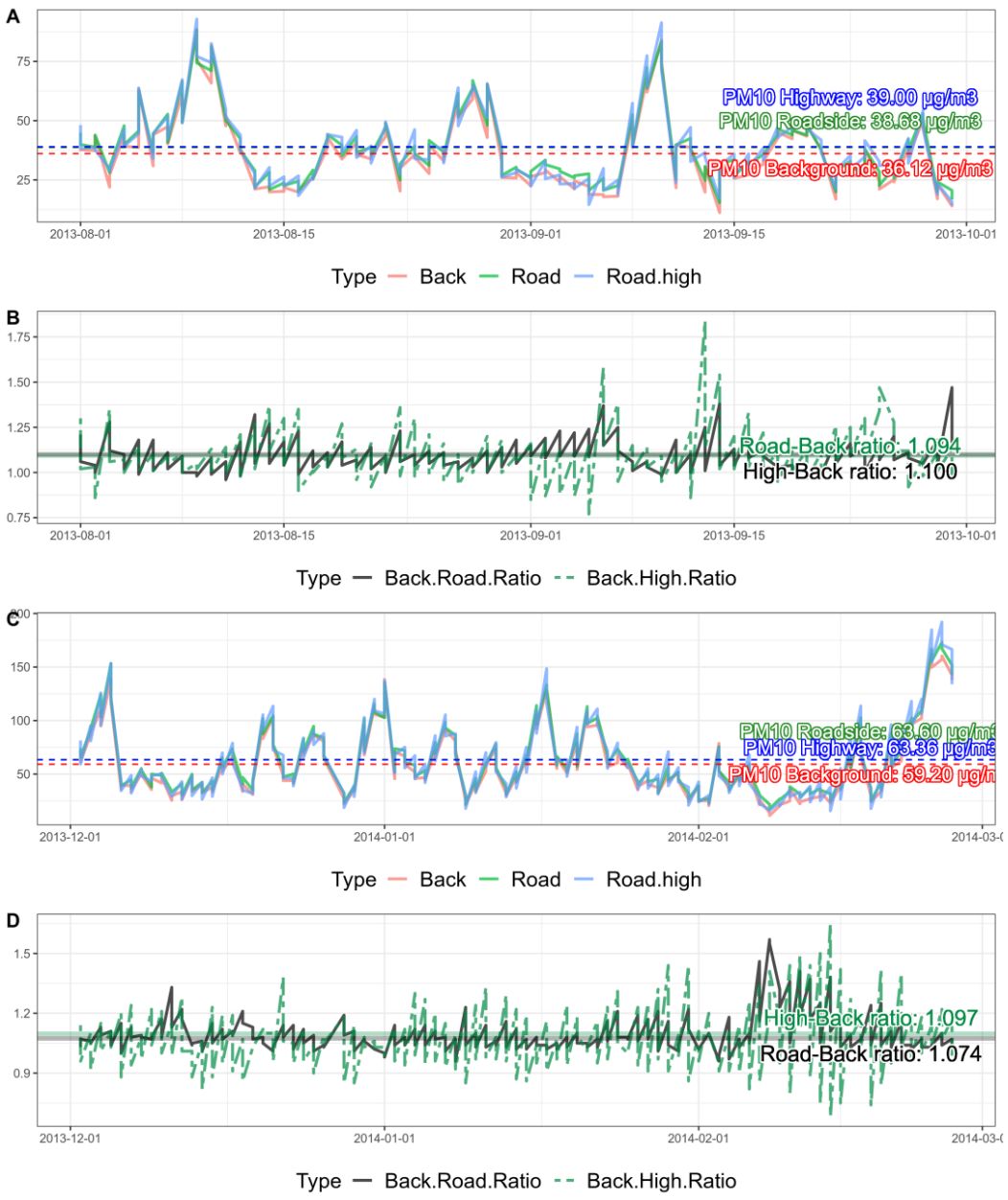


Fig. 3.20 PM_{10} ratios measured by every 12-hour average in late July-September 2013 and December-February 2013-2014. Graphs A and C are pollution levels of PM_{10}, and Graphs B and D are ratios between aggregated roadside and background stations, and between aggregated urban highway and background stations

3.4 Results

3.4.1 Comparison of Interpolation Results

This section compares NO_2 and PM_{10} maps computed from UK and GAM, along with the road-overlayed results during mid-August 2013 and mid-February 2014 (see from Figure 3.21 to Figure 3.26; for the full results please refer to the Appendix).

In Figure 3.21, the NO_2 concentration in mid-August was fairly low at an average of 19.39ppb (min: 13.6ppb, max: 31.1ppb). The distribution was homogeneous across the city except for the last two days where the concentration became higher in the city centre and the southeastern region. The road-overlayed results increased to 27.87ppb, but with the road to non-road ratio the outcomes ranged from 9ppb to 134ppb, with the highest being on the 20th. In comparison, NO_2 in mid-February had a relatively higher average of 34.5ppb with the lowest at 9ppb and the highest at 58ppb (see Figure 3.22). February 20th had the highest concentration during the entire period where the mean concentrations exceeded 40ppb (±sd 6.4). The road-overlayed results showed a marginal increase to the overall concentration at 42ppb, but the places nearer to road layers exceeded over 100ppb on the 15-17th and the 20th. From Figure 3.21(c) and Figure 3.22(c), the difference between methods was overall small in both seasons (mostly yellow), and even days when the difference is higher than 1ppb, the difference rate lies below 10 per cent which is unlikely to give large health effects.

For PM_{10}, the concentration in mid-August ranged from 19µg/m^3 to 50µg/m^3, where the concentration was, at most times, distributed homogeneously (see Figure 3.23). August 11th was the highest day amongst the maps which exceeded the EU limit of 40µg/m^3, but was less than Korea's daily limit of 100µg/m^3. Although the average PM_{10} increased a small amount to 43µg/m^3 in mid-February, there were exceptionally polluted days - 16th, 17th, and 20th - when it became closer to the nation's legal limit. With the road weighted results, which are assumed to be more realistic, there was a higher average at 47µg/m^3; however the concentration nearer to the road peaked over 100µg/m^3 on the - 16th, 17th, and 20th - with the highest at 175µg/m^3 in the daytime hours of the 17th. The difference between methods was mostly small and they have a similar structure, having on average less than 1µg/m^3 difference. In mid-February, the trends were also similar between methods except on the 17th-19th. The difference maps (see Figure 3.23c and Figure 3.24c) also illustrate that there was almost no difference between the two methods, and even on days, such as February 18th, where GAM is underfitted and UK is overfitted, the concentration only differs by 6% which is not concerning.

In the road-overlayed results (see Figure 3.25 and Figure 3.26), NO_2 had a discrete difference to non-road areas while PM_{10} showed a marginal difference. Based on the ratio calculation (see Section 3.3.6), the NO_2 on roads was consistently higher than the background areas, mainly due to the volume of traffic that generates and disperses

tailpipe gases to nearer roads. NO_2 quickly dilutes when the wind blows hard. PM_{10}, is also a component generated from tailpipe emission and when tyres elevate dust but also is affected by heating weather conditions. However, PM_{10} will settle out in a way that NO_2 does not. For example, if the ground is cold and there is little wind the air mass will be trapped and cause a higher pollution episode.

Although recent studies found less association between health risks and living close to major roads due to the time of staying in the high-polluted areas and the difficulty of including their medical history (Beelen et al., 2014; Heinrich et al., 2013), this approach would help understanding the adverse effects from traffic intensity and road proximity.

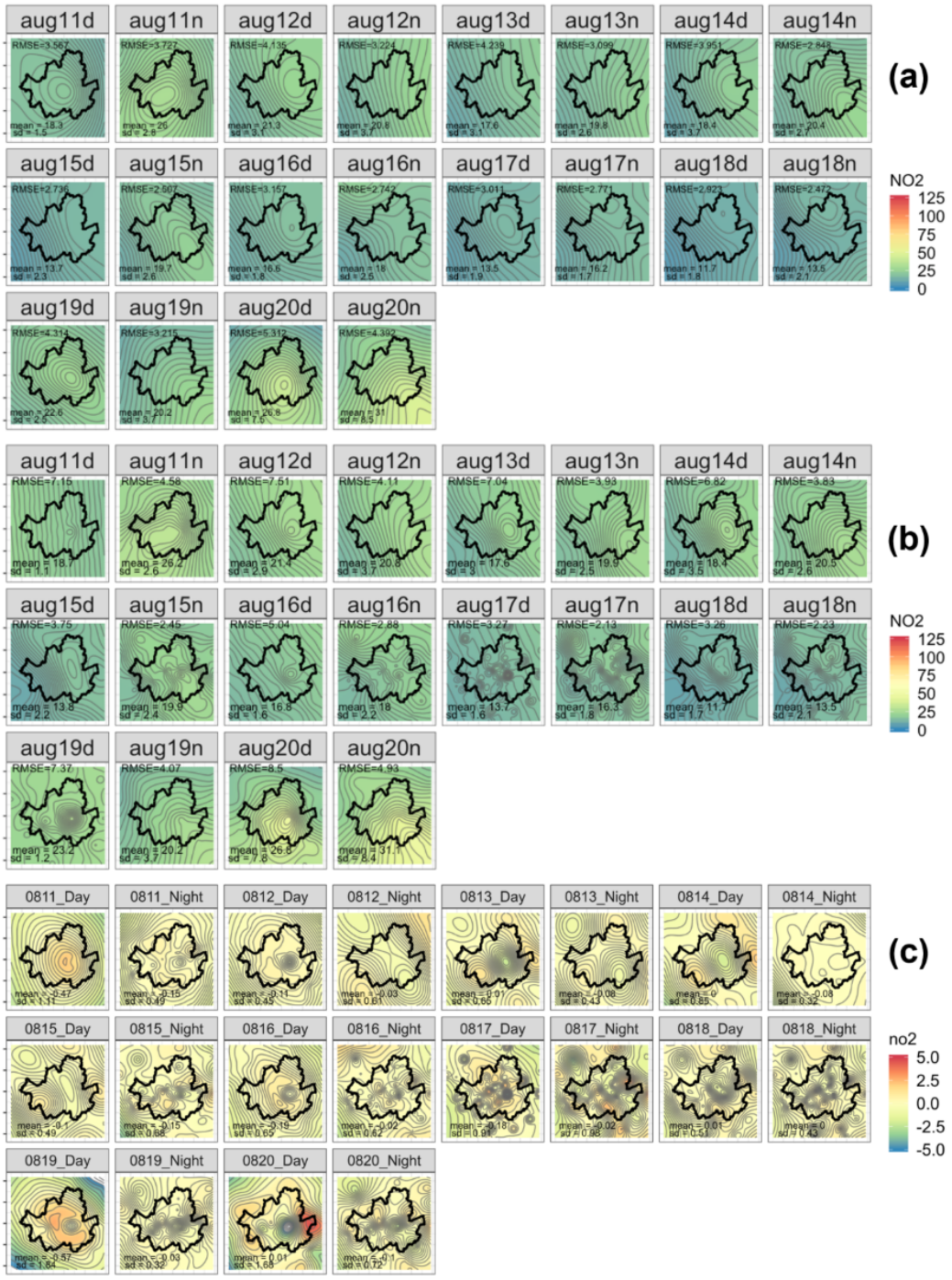


Fig. 3.21 Interpolated NO_2 maps on August 11-20th. Maps (a) and (d) are the outcomes of GAM; (b) and (e) are the outcomes of UK; (c) and (f) are the difference outcomes between methods (c=a-b, f=d-e). Units:ppb

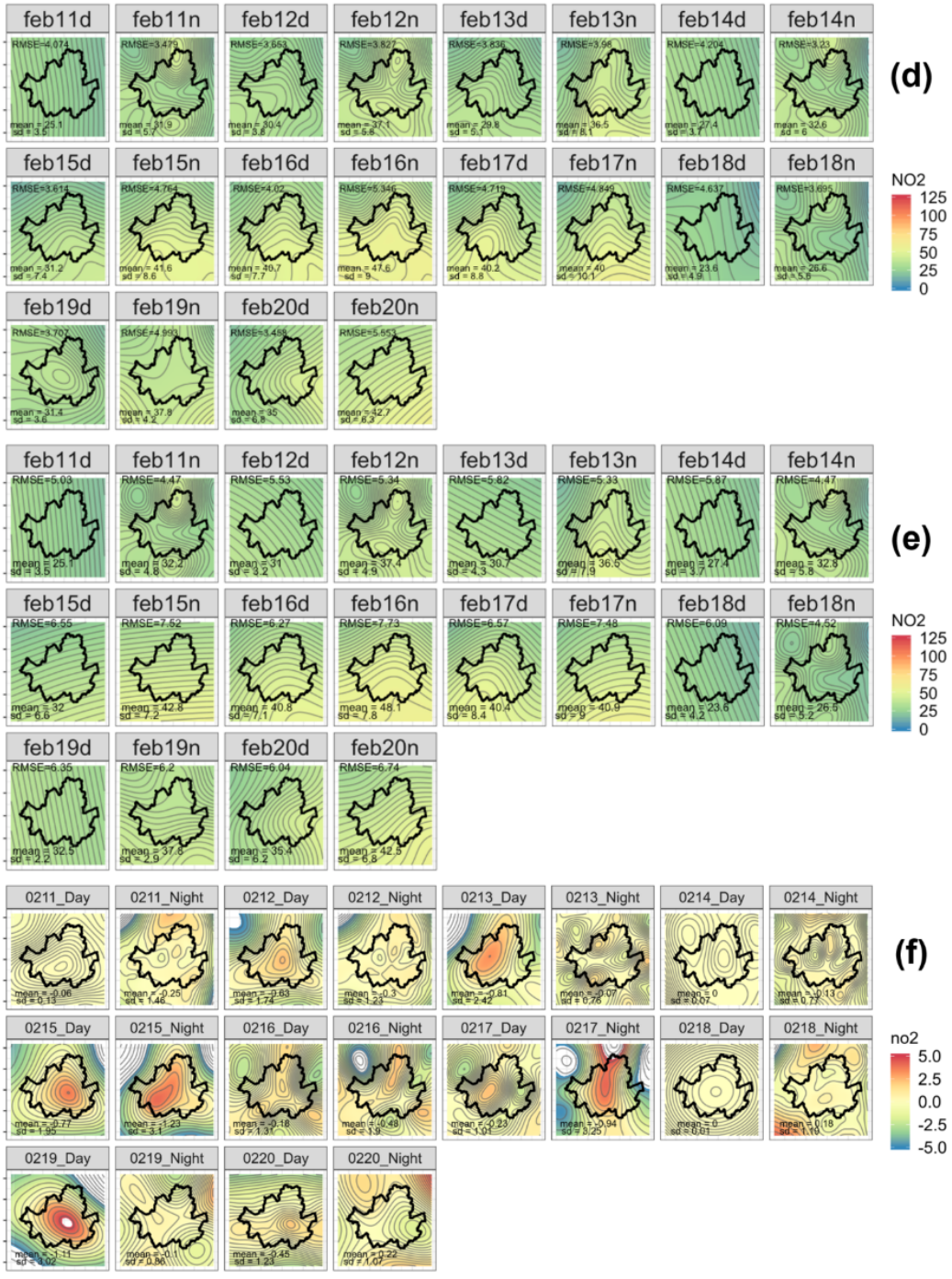


Fig. 3.22 Interpolated NO_2 maps from August 11th-20th. Maps (a) and (d) are the outcomes of GAM; (b) and (e) are the outcomes of UK; (c) and (f) are the difference outcomes between methods (c=a-b, f=d-e). Units:ppb

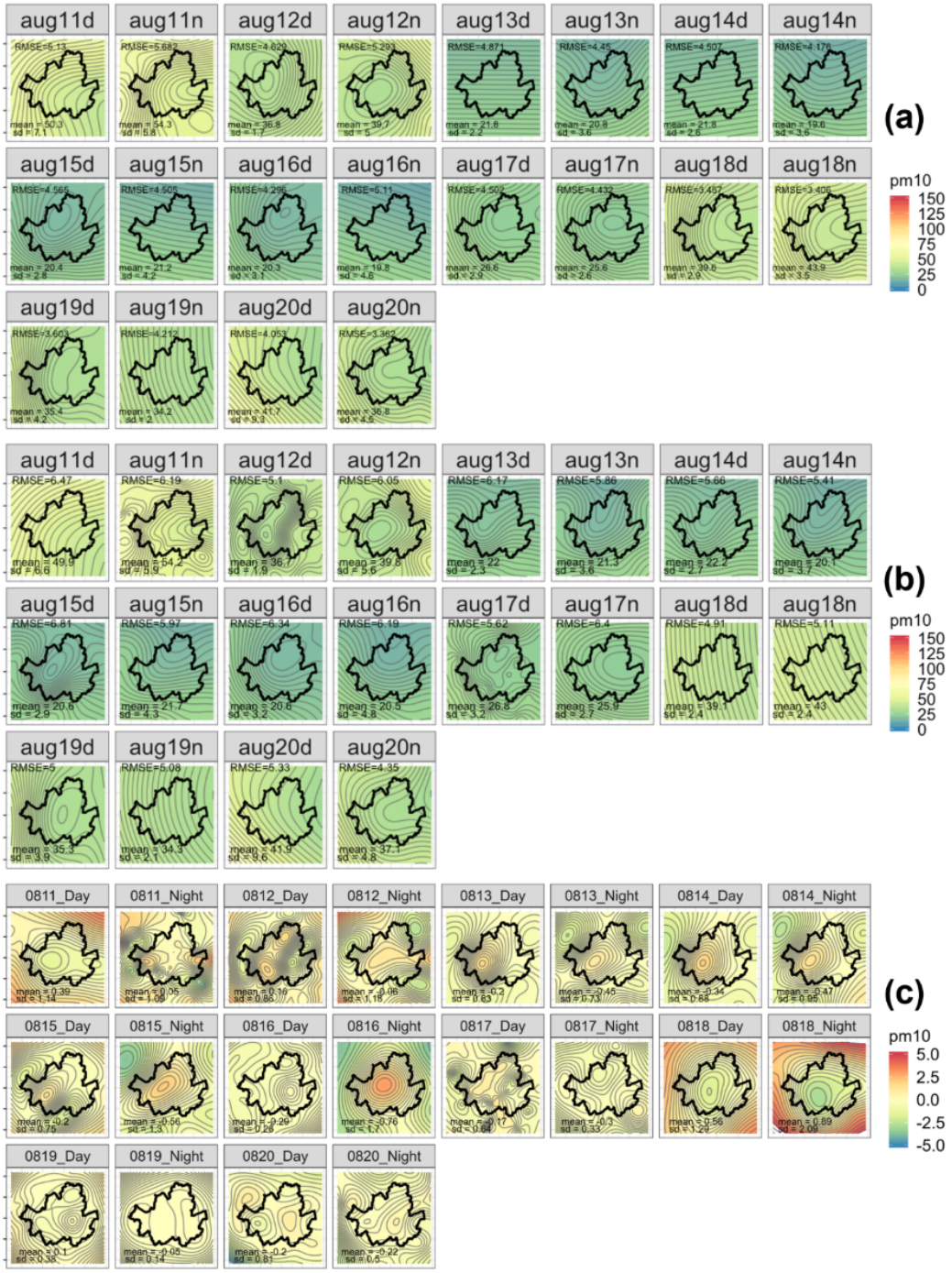


Fig. 3.23 Interpolated PM_{10} maps from August 11th-20th. Maps (a) and (d) are the outcomes of GAM; (b) and (e) are the outcomes of UK; (c) and (f) are the difference outcomes between methods (c=a-b, f=d-e). Units: $\mu g/m^3$

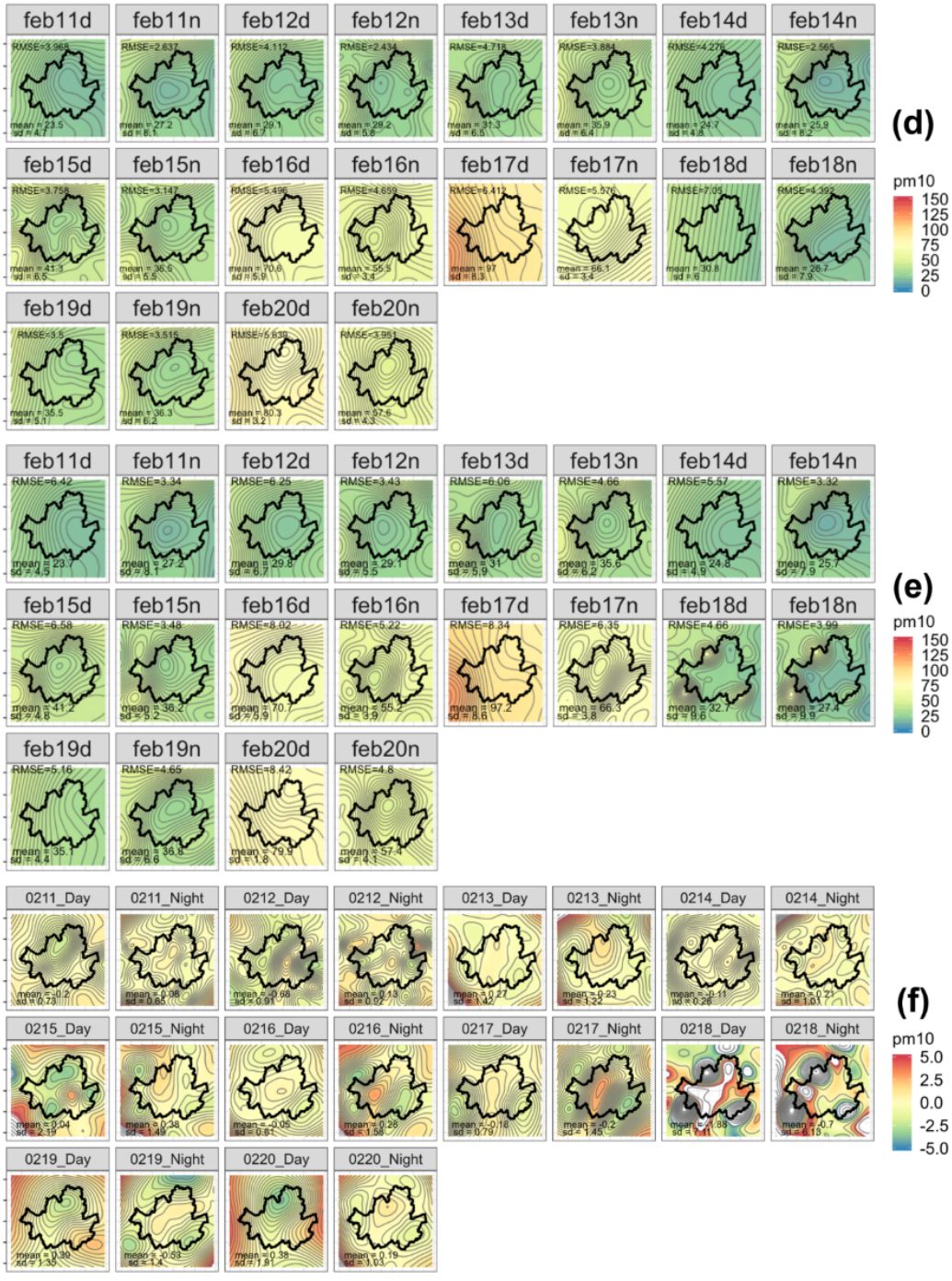


Fig. 3.24 Interpolated PM_{10} maps from February 11th-20th. Maps (a) and (d) are the outcomes of GAM; (b) and (e) are the outcomes of UK; (c) and (f) are the difference outcomes between methods (c=a-b, f=d-e). Units: $\mu g/m^3$

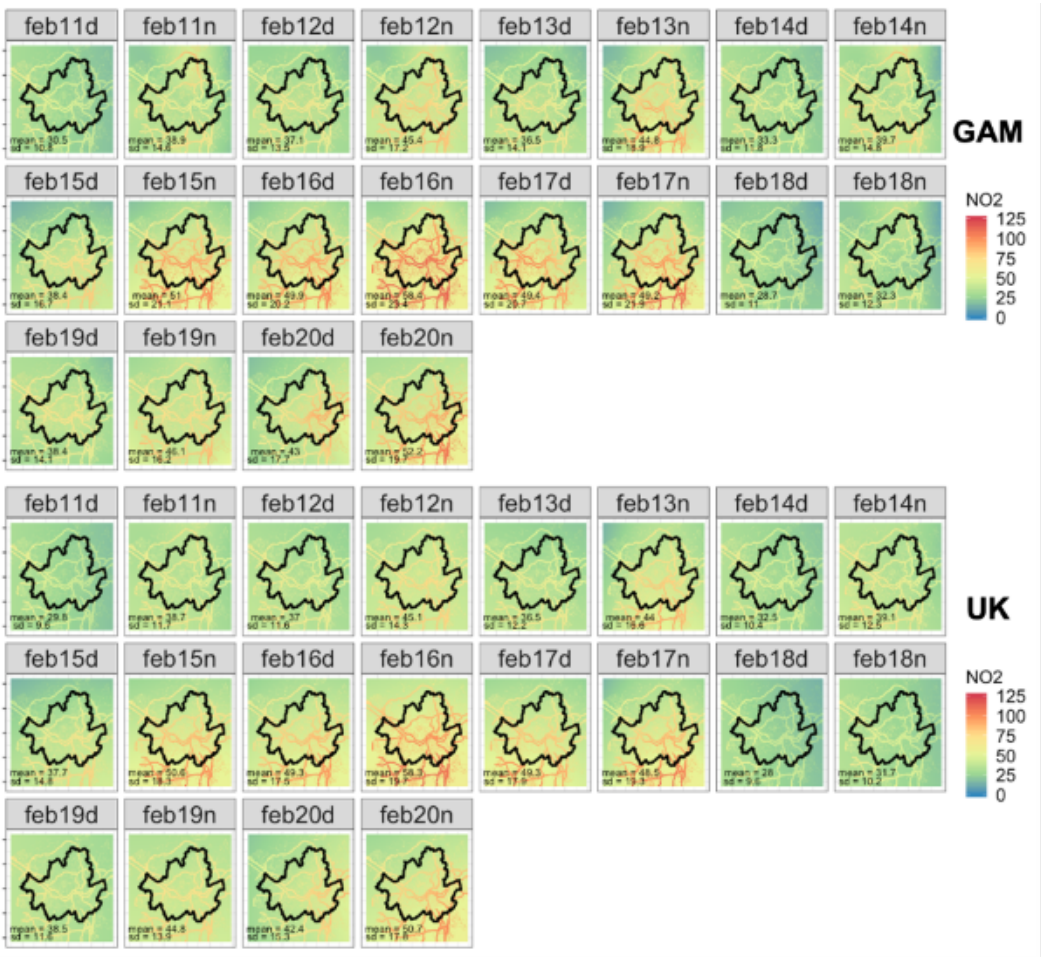


Fig. 3.25 Time series outcomes of combined NO_2 interpolation and road effects in mid-February 2014. Units: ppb

Fig. 3.26 Time series outcomes of combined PM_{10} interpolation and road effects in mid-February 2014. Units: $\mu g/m^3$

3.4.2 Temporal variability of SI results

Figure 3.27 and Figure 3.28 show a half-day aggregation of NO_2 interpolation with GAM and UK in August, September, and December 2013, as well as January and February 2014. The aggregation was made over the whole region.

The distributions displayed for NO_2 had an oscillation of peaks and troughs every 3-4 days, and the concentration was explicitly higher during the winter period. NO_2 levels varied between 10ppb and 45ppb in summer, with a seasonal average of 29ppb. An interesting fact was that boxplots with wider ranges tended to appear when an upward trend was observed. The 3-day oscillation pattern continued in winter but with higher concentration ranging from 5-94ppb. Although the average NO_2 levels complied with the national standards, there were over 10 observations (half days) where NO_2 exceeded the legal limit. Also, winter had less NO_2 variation which means that the pollutant was more uniformly distributed over the urban area.

Figure 3.29 and Figure 3.30 show a half-day aggregation of PM_{10} interpolation with GAM and UK in the same period as the NO_2 results, with the horizontal line as the monthly average. There were three occasions in summer - August 9-10th, and September 10th - that almost reached the daily limit of 100µg/m^3, but winter had worse 30 occasions (15 days) which actually exceeded the daily limit, the last week of February exceeding it constantly day and night. There weren't any oscillation patterns discovered in PM_{10}, but PM_{10} tends to maintain the concentration for a while, then instantly peaks for a short occasion, then decreases again. This tendency becomes more frequent during winter.

This study discovered that the seasonal patterns of summer trough and winter peak appeared in both pollutants, whereas the daily oscillation was only found in NO_2 results. This result was similar to the annual air quality report from the environmental government of Seoul where they described the seasonal cycles of NO_2 and PM_{10} and attributed them to meteorological conditions (wind, atmospheric stability, precipitation), car use and domestic heating, and daily cycles of NO_2 due to photochemistry effects and the frequency of car use (Roberts–Semple et al., 2012; Song et al., 2011).

In terms of performance comparison, the medians, ranges and extremes were rather similar overall. The average for GAM was marginally higher than the result for UK on some days, but because of the difference in model setting, and a semivariogram against knot configuration, the difference in units was small.

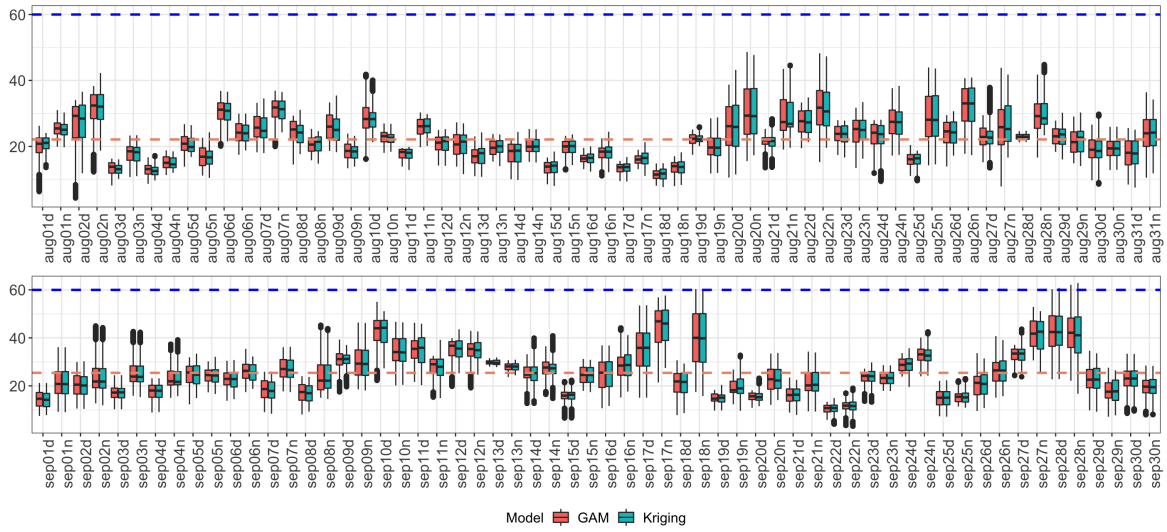


Fig. 3.27 Temporal Boxplot of NO_2 in Summer 2013. The blue line indicates a daily national standard (60ppb) and the red line indicates the seasonal average. Units: ppb

3.4.3 Spatial Variability Between SI Results

Although the boxplot analysis provided the temporal variation of pollution, it did not capture the variability in space which is important when estimating population exposure by location.

Figure 3.31 and Figure 3.32 show the aggregation of NO_2 and PM_{10} in the first, middle and last days of each month. Each map visualises the variance from the average. For example, if a location has a value of -5, it would mean that the pollution level is -5 from the mean of the map.

For NO_2, higher concentrations were either discovered from the central or the southeastern parts of Seoul for the entire period that ranged around 4-15ppb higher than the average concentration, whereas the northern areas were relatively lower. The difference between methods was less than 1ppb on average, although the contour patterns were slightly different. For example, on the August 11th-20th maps, the highest two contours were almost symmetrical, but the kriging outcome had more local effects to take into account. A similar result was observed on December 21st-31st.

The most noticeable difference was in the first 10 days of August. In Figure 3.31, the first two columns of September showed that GAM had multiple ring ellipses of concentration heading in the NW-SE direction while UK had smaller ellipses and a

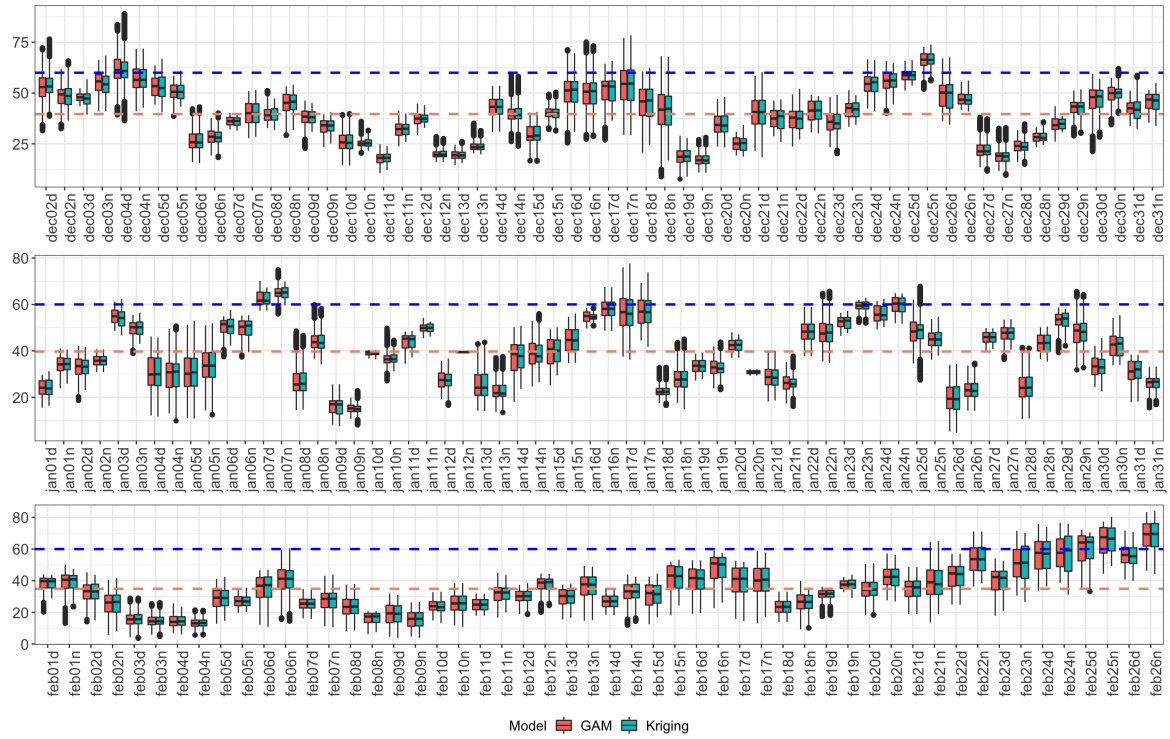


Fig. 3.28 Temporal Boxplot of NO_2 in Winter 2013-2014. The blue line indicates a daily national standard (60ppb) and the red line indicates the seasonal average. Units: ppb

few high local values. The grey-shaded legend also distinguishes the two methods by highlighting a 'dip' in the UK method whereas GAM has a smoother outcome.

In contrast to NO_2, a lower PM_{10} was captured at the central and eastern parts of the city (see Figure 3.33 and Figure 3.34). The authors of of Seoul (2016) argued that the PM_{10} increase was highly correlated with air pressure, which normally brings pollution from the westerly and southerly winds where the industrial areas are situated, which supports the fact that the western areas are highly polluted. There is a controversial argument about whether the Chinese industrial complex, built on the east coast of China (around Shandong peninsula), triggered the increase of PM_{10} in Korea.

There were days where high concentrations of PM_{10} was distributed in the north and west while the centre and to the south was low. These days include the last 10 days of December and January. These changes might have been affected by the northerly wind that carried dust from the north industrial areas as well as the arid lands near the North Korean border.

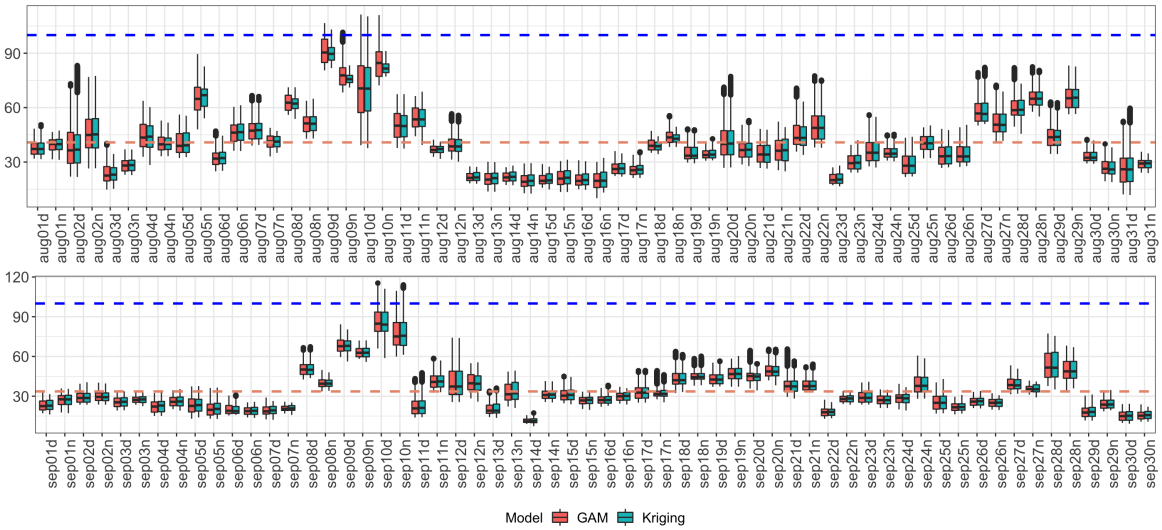


Fig. 3.29 Temporal Boxplot of PM_{10} in Summer 2013. The blue line indicates a daily national standard (100µg/m^3) and the red line indicates the seasonal average. Units: µg/m^3

The findings from the spatial variability convey some useful insights. First, the relative values are easily comparable in terms of Seoul's geographical distribution – low in the northwest and northeast, high in central Seoul and the southeast. Instead of using an absolute concentration, it immediately allows us to think more about the landscape and built environments that affect the difference. Second, it is easy to compare the outcome of GAM and UK, although the general picture looks quite similar.

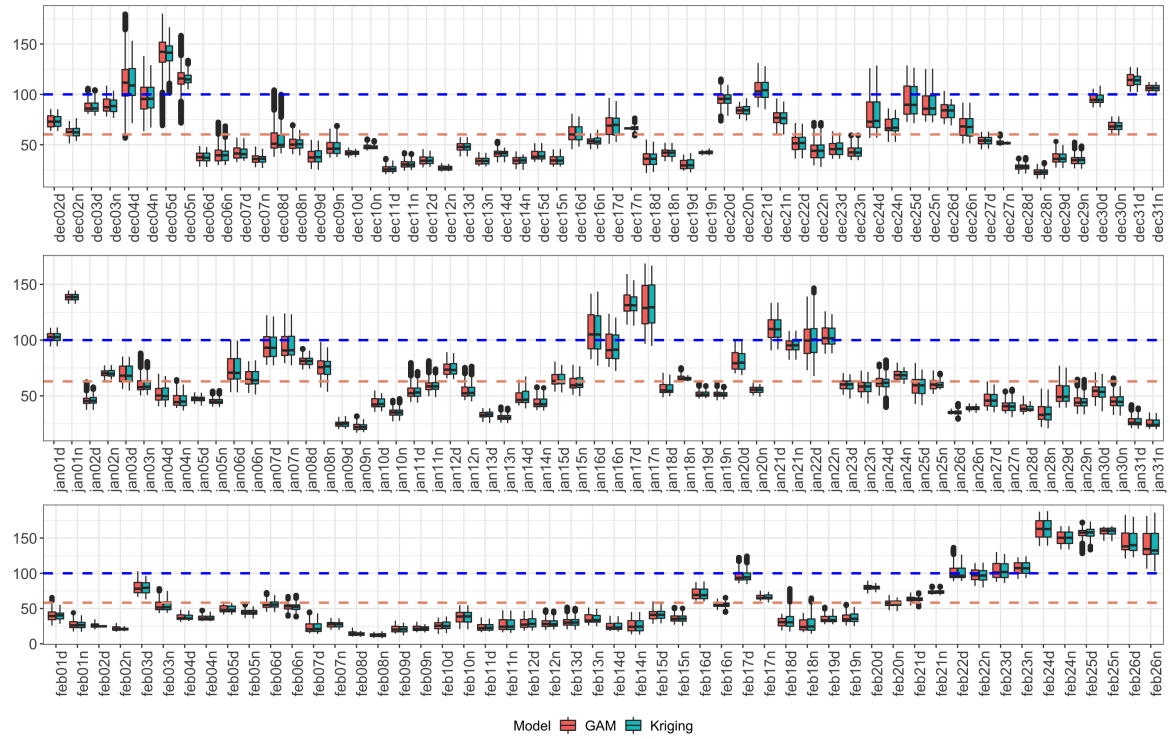


Fig. 3.30 Temporal Boxplot of PM_{10} in Winter 2013-2014. The blue line indicates a daily national standard (100µg/m^3) and the red line indicates the seasonal average. Units: µg/m^3

3.4.4 Validation: Scatter Plots

One way of testing RMSE (Root Mean Square Error) is to use scatter plots. This section implemented four separate regressions between the modelled results and observations: GAM-observation, UK-observation, GAM final-observation, and UK final-observation. The regression model was applied to groups at a 10-day time step to capture RMSE. Figure 3.35 to Figure 3.39 show the NO_2 results of the 1st-10th of August, September, December, January, and February in 2013-2014.

The results showed that while the modelled GAM and UK had a significant correlation, some stations that are close to roads were overly corrected in systematically. In summer, GAM-observation and UK-observation had a reasonable fit with similar RMSE values, but the results from GAM final-observation and UK final-observation had a worse correlation (Figure 3.35 and Figure 3.36). This was because the stations nearer to roads overly corrected the results. The outliers were from Jungnang.

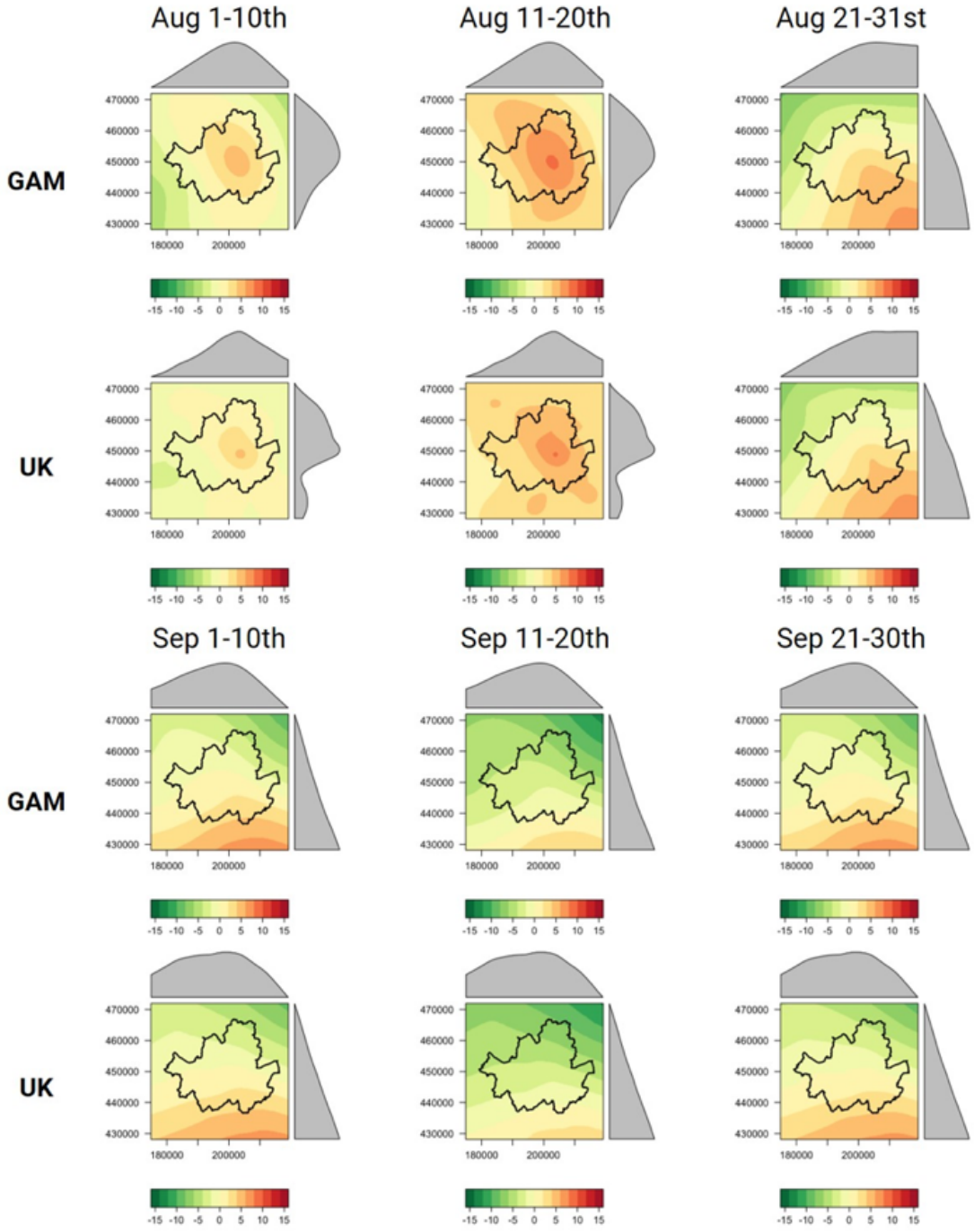


Fig. 3.31 Spatial variability of NO_2 in Summer 2013 (Units: ppb)

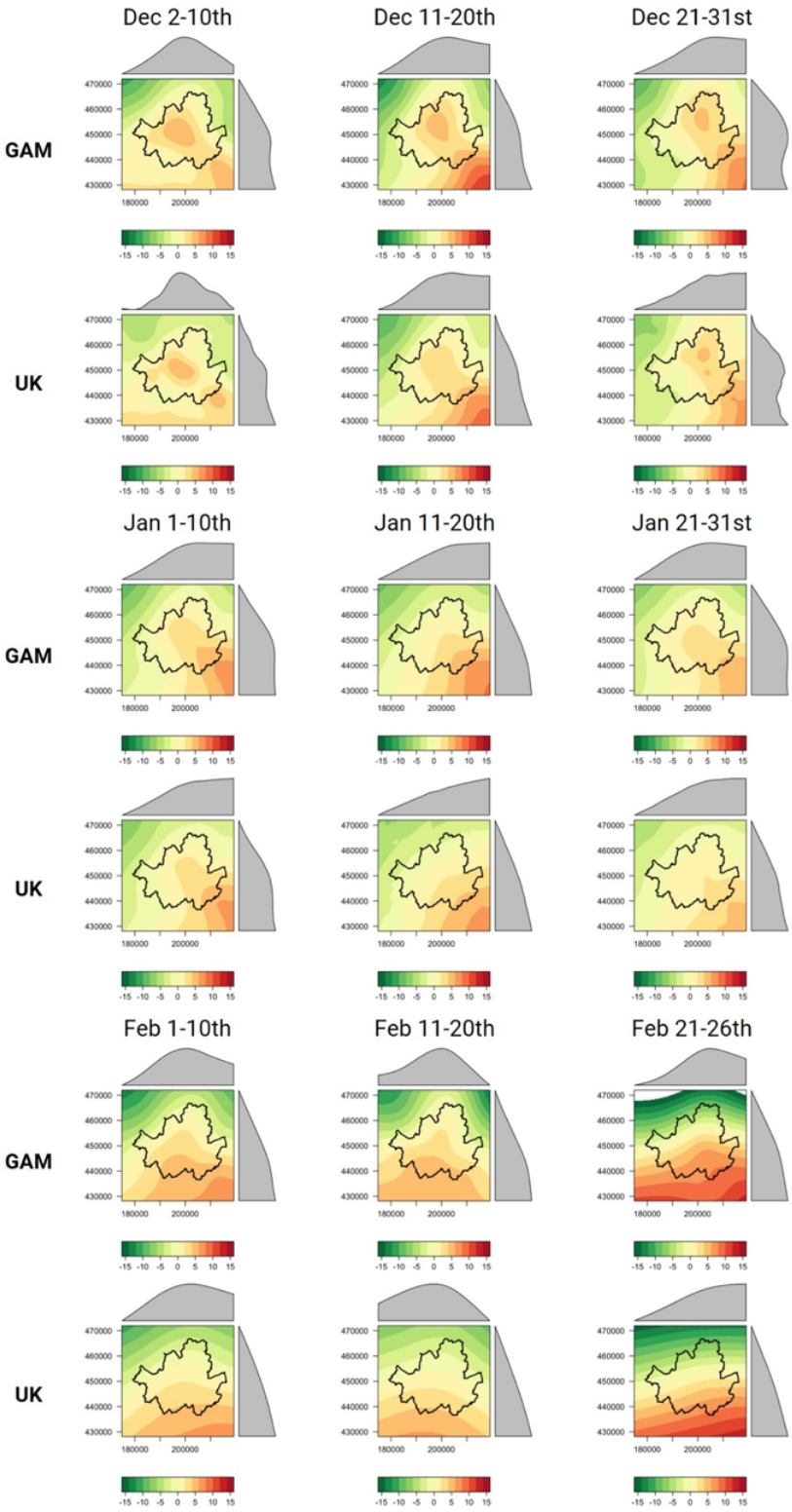


Fig. 3.32 Spatial variability of NO_2 in Winter 2013-2014 (Units: ppb)

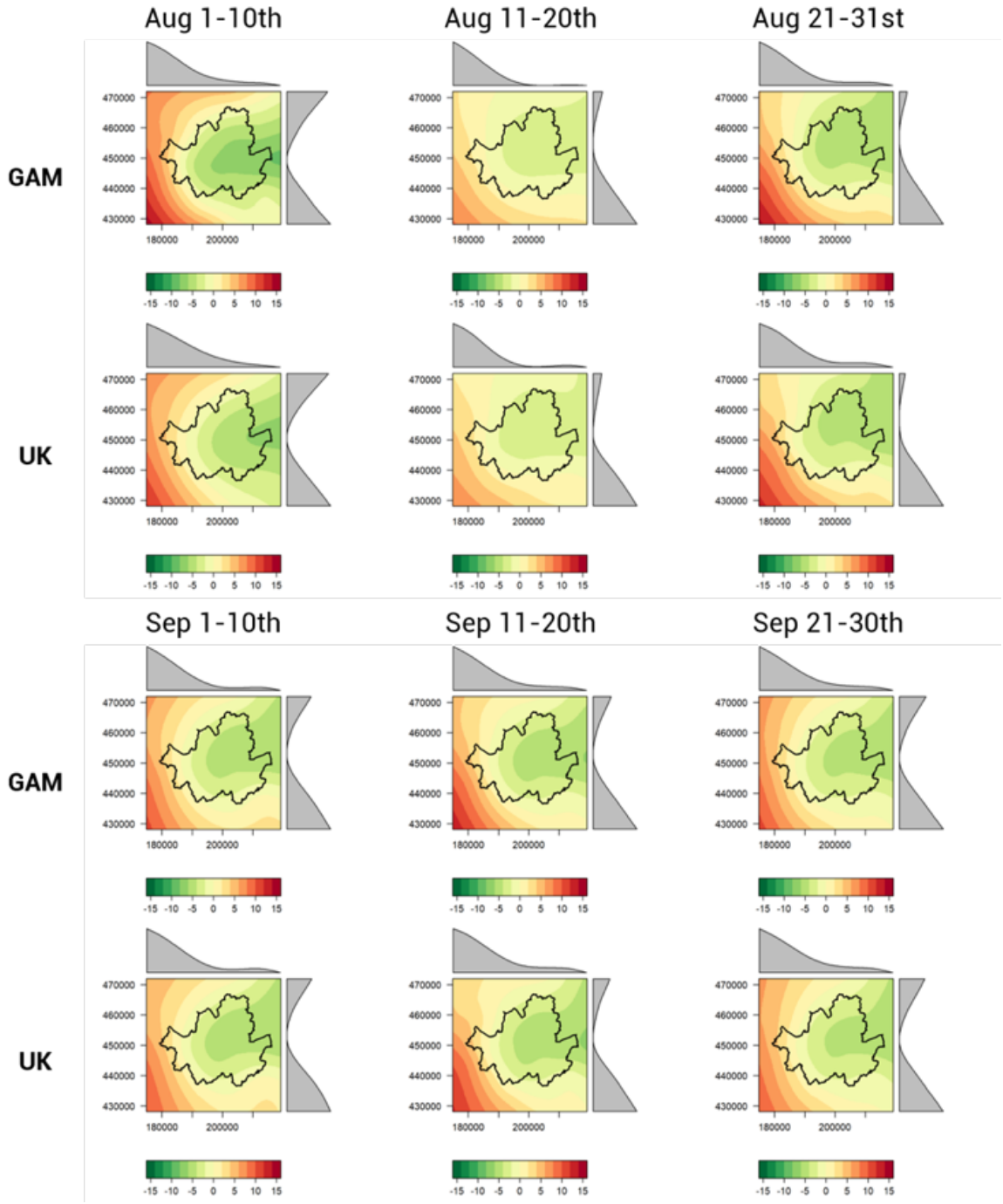


Fig. 3.33 Spatial variability of PM_{10} in Summer 2013 (Units: $\mu g/m^3$)

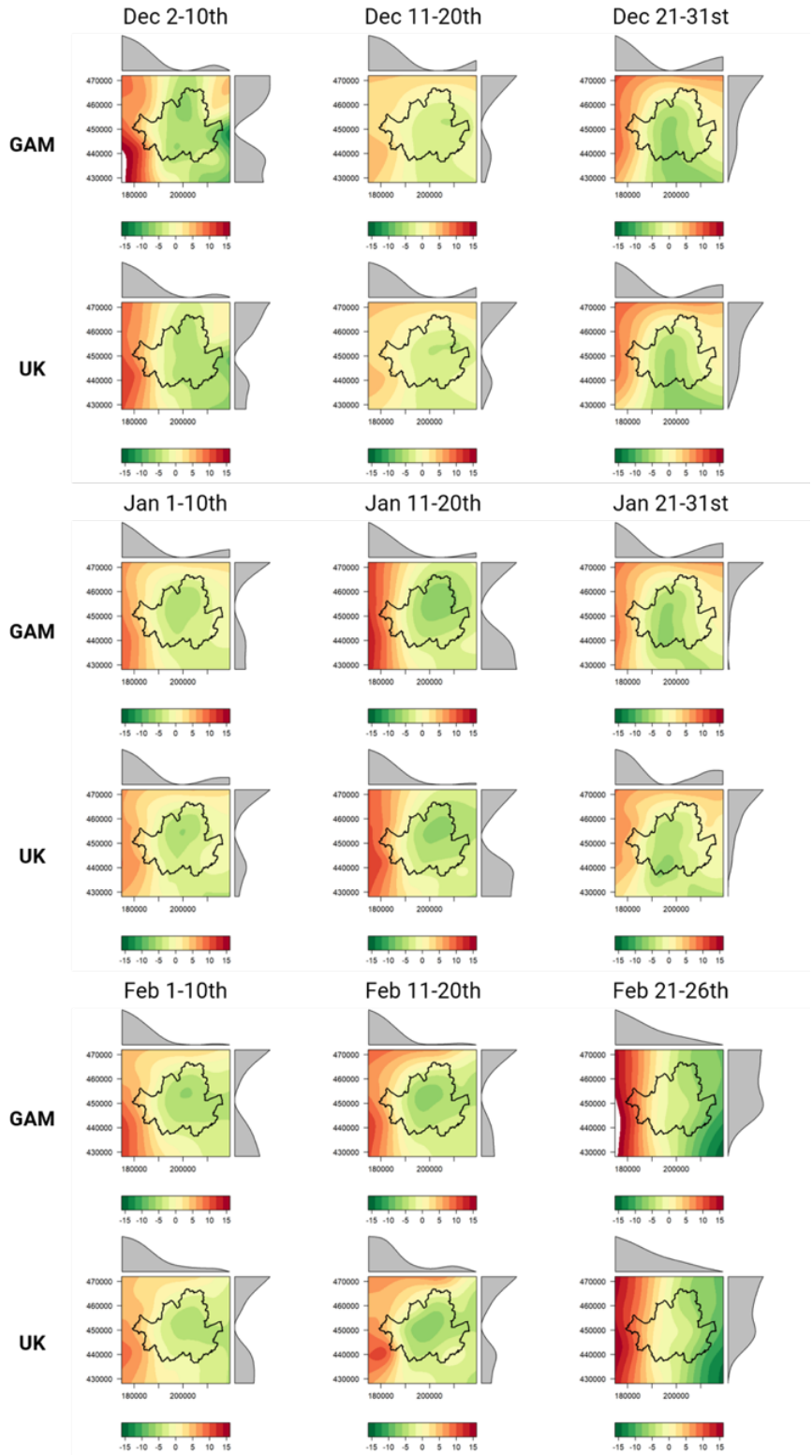


Fig. 3.34 Spatial variability of PM_{10} in Winter 2013-2014 (Units: µg/m^3)

Table 3.5 RMSE Comparison of NO_2 and PM_{10} between GAM and UK in August and September 2013

Pollutant	Method	August			September		
		Top	Mid	Bot	Top	Mid	Bot
NO_2	GAM	4.34	3.07	4	4.05	3.77	3.39
NO_2	UK	3.62	2.38	3.56	3.83	3.29	2.82
PM_{10}	GAM	4.17	4.1	3.62	3.62	3.56	3.2
PM_{10}	UK	4.74	3.95	3.46	3.68	3.29	3.08

Table 3.6 RMSE Comparison of NO_2 and PM_{10} between GAM and UK in August and September 2013

Pollutant	Method	December			January			February		
		Top	Mid	Bot	Top	Mid	Bot	Top	Mid	Bot
NO_2	GAM	4.38	3.92	4.99	4.41	4.5	4.59	3.28	3.85	5.22
NO_2	UK	3.88	3.8	4.42	4.31	4.45	4.49	3.48	4.11	5
PM_{10}	GAM	6.09	4.48	5.12	5.11	4.95	4.17	3.31	3.95	8.08
PM_{10}	UK	6.63	4.4	5.01	4.73	4.81	3.49	3.46	3.53	8.02

Compared to summer, the modelled outcomes were also highly correlated in winter (see Figure 3.37, Figure 3.38, and Figure 3.39). UK had a lower RMSE than GAM in December and January but the difference was less than 0.05. The significance in February was slightly lower than the GAM. The overall performance of the GAM and UK with roads did make an improvement but the degree was not as striking as summer, i.e. summer made more difference than winter.

3.5 Discussion and Conclusion

This study explored various pollution sources on a temporal domain and compared two spatial interpolation methods and applied an extra road effect to estimate a spatially gridded field as well as possible given the sparse spatial data.

Chemical and Health Aspects

For the data analytics section, this study used NO_2 and PM_{10} to understand different temporal patterns based on the tidyverse package in R (Wickham, 2019). Understanding

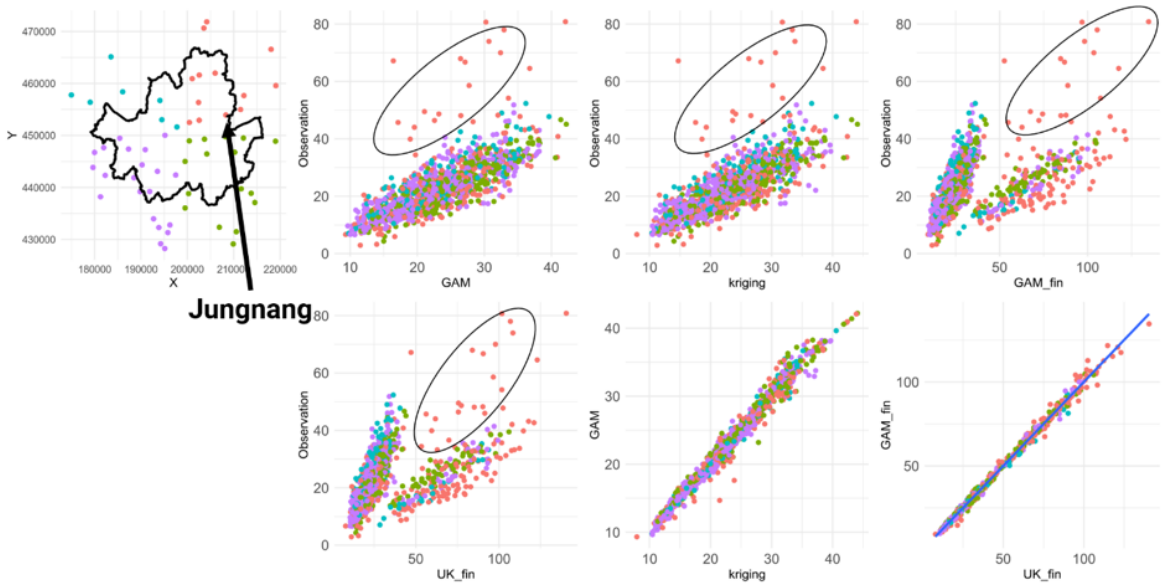


Fig. 3.35 NO_2 validation results of six different variables from August 1st-10th 2013. The map on the top left shows the distribution of stations surrounding Seoul, with four different colour schemes from the centroid (NW, NE, SW, SE). GAM-observation and UK-observation had a reasonable fit, but the stations nearer to roads gave rise to the modelled results that led to overprediction. The outliers (circled area) were from Jungnang. Note that a point means a half-day of a station, thus a single station repeats 20 times (this is why all the circled areas are from one station).

the data is the foundational and crucial work to build the rest of the study. The results showed that the gaseous pollutants (NO_2) presented an hourly, daily, and seasonal cycle, while PM_{10} was less sensitive to the daily cycle but followed an annual oscillation.

From the time-series analysis, a decreasing trend was discovered from all pollutants, particularly after 2014. This might be affected by the new pollution mitigation plan by the government after the WHO announced the severe effect of air pollution in 2013 (WHO, 2012).

From the daily perspective, NO_2 had a distinctive cycle within a day, with a strong inverse relationship between night and day. NO_2 peaked twice a day, of which the first was at 07-10h and the second at 18-22h. The findings extend the previous studies Song et al. (2011) and Roberts–Semple et al. (2012), confirming that the NO_2 concentrations spike during rush hours due to the basic turbulence of pollutants led by surface heating from the sunlight, but mostly from heavy road traffic. Huang et al. (2001) elucidated the double-peak patterns of NO_2 by reasoning that the photochemical disappearance

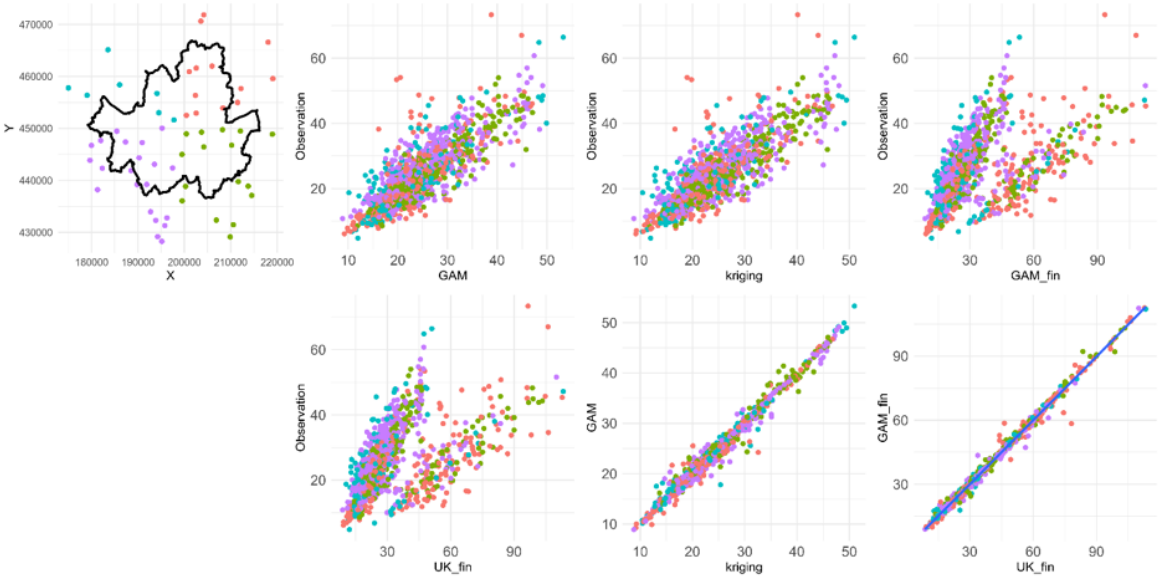


Fig. 3.36 NO_2 validation results of six different variables from September 1st-10th 2013. The map on the top left shows the distribution of stations surrounding Seoul, with four different colour schemes from the centroid (NW, NE, SW, SE). GAM-observation and UK-observation had a reasonable fit, but the stations nearer to roads gave rise to the modelled results that led to overprediction.

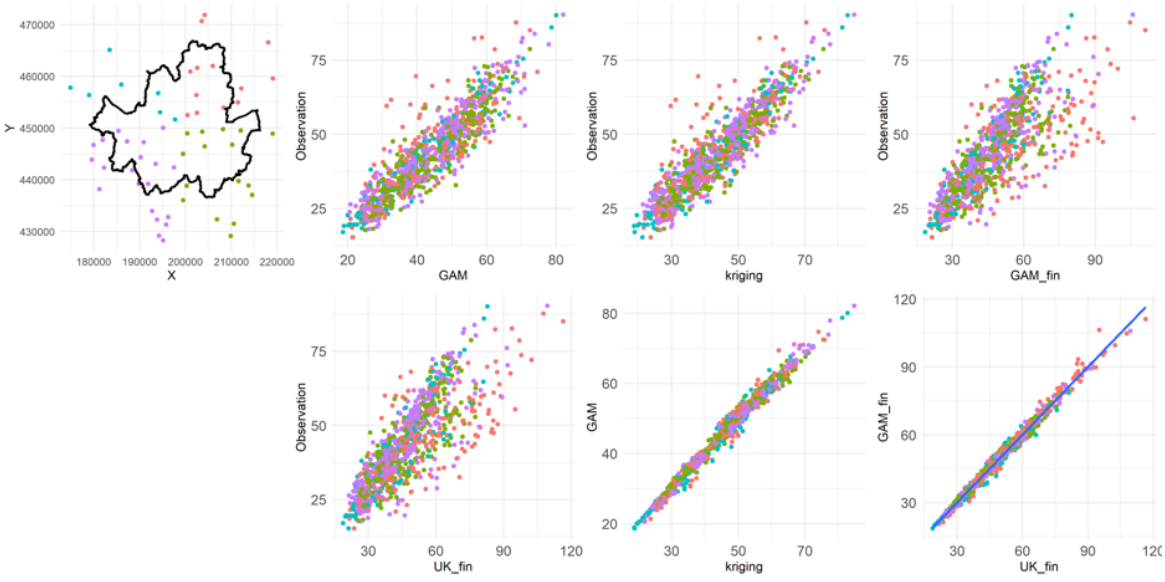


Fig. 3.37 NO_2 validation results of six different variables from December 1st-10th 2013. The map on the top left shows the distribution of stations surrounding Seoul, with four different colour schemes from the centroid (NW, NE, SW, SE).

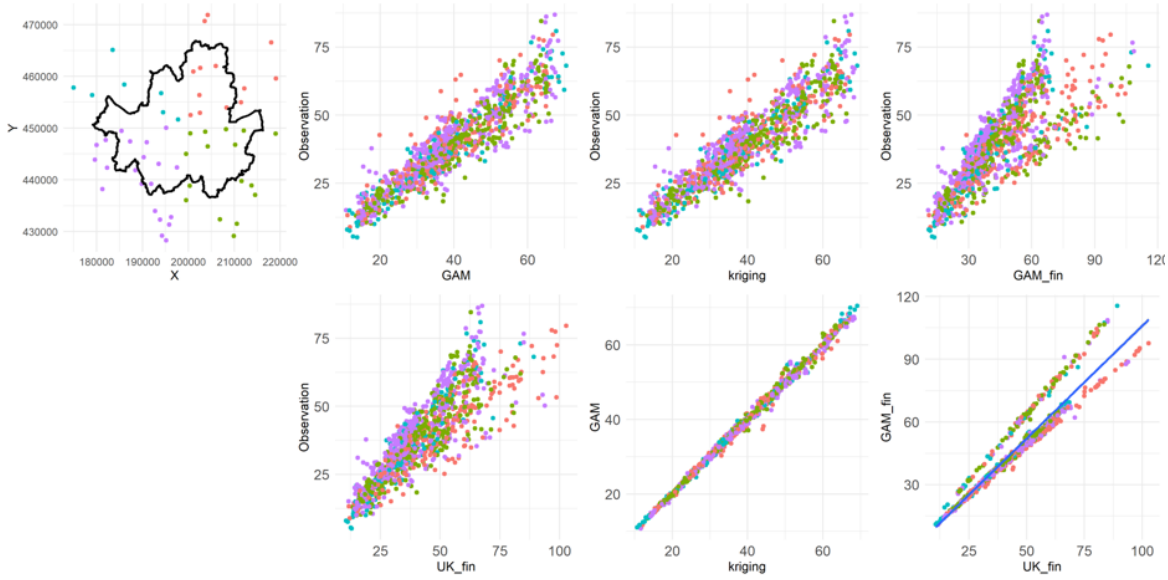


Fig. 3.38 NO_2 validation results of six different variables from January 1st-10th 2014. The map on the top left shows the distribution of stations surrounding Seoul, with four different colour schemes from the centroid (NW, NE, SW, SE).

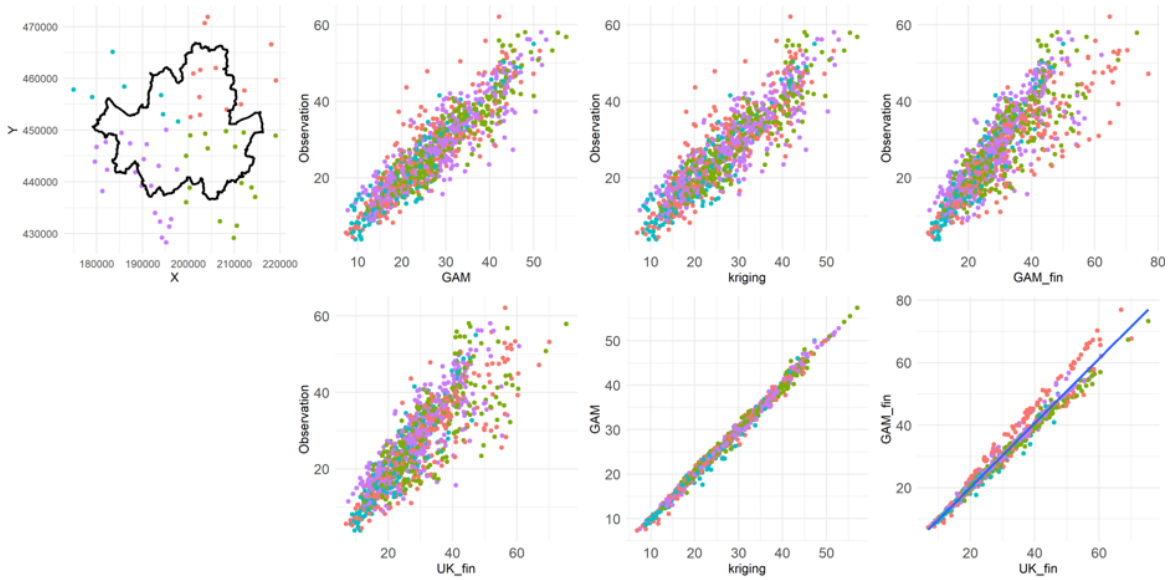


Fig. 3.39 NO_2 validation results of six different variables from February 1st-10th 2014. The map on the top left shows the distribution of stations surrounding Seoul, with four different colour schemes from the centroid (NW, NE, SW, SE).

of O3 after sunset and added that the prolonged hours of the second peak was from a spontaneous reduction in emissions. However, the author argues that a second peak is

not just the matter of photochemistry but also from another peak of traffic volume that moves out of the working districts (Roberts–Semple et al., 2012), but at a dispersed time. Late-night public transportation in Seoul can also be a factor.

From a health aspect, the decrease of NO_2 is not always reassuring because the sunlight converts NO_2 vehicle exhaust fumes to dangerous ground-level O_3 (DEFRA, 2003). This implies that pedestrians who have lung diseases or are otherwise vulnerable must, therefore, be advised to stay indoors or be aware of walking on that day. Results between background and road stations also support the idea that some pollutants may be more from a risk near roads. Concentrations of NO_2 were at least 1.5 times higher than the background areas, while PM_{10} showed less than a 10% difference. But the more important fact is that the roadside results from NO_2 (42-50ppb) exceeded the annual legal limit of 30ppb, which in turn is a warning to people who walk near roads or sit in vehicles that they have more risk to their health (Schraufnagel et al., 2019). The concentration of PM_{10} (47-55ppb) and $PM_{2.5}$ (25-27ppb) suggests that the small difference between station types explains that the particles were well mixed in the air Keuken et al. (2012); however, is still a matter due to the exceedance of the legal limit.

Regarding the results from GAM and UK, it was found that with the NO_2 there were oscillations every 3 days, but for some of the winter days, it did not comply with the daily limit of 60ppb. The results align with the findings from (Huang et al., 2001), where the NO_x level in winter was higher than summer due to the later sunrise and shorter sunlight hours that weaken the photochemical reaction. From the temporal results, the NO_2 variation of daily concentrations was greater in summer, notably when the concentrations started to rise (Aug 19th-21st, Aug 25th-27th, Sep 16th-18th, and Sep 27th-28th). For these days, it is estimated that the NO_2 level was perhaps concentrated in particular areas such as CBD, subcivic zones, or highways. Performance-wise, it was difficult to find the difference from either method, with a maximum of 3ppb.

Results of spatial variability inform us that the northwestern and northeastern areas had consistently low NO_2 whereas the central and the southeastern areas had consistently high NO_2. High traffic volume can directly explain the surpassing levels of NO_2 for central Seoul, regardless of the absolute concentration, and for the southeastern high, it might have resulted from the mixture of flows of major highways that lead to Seoul and a massive local network that was constructed for the new towns and techno valley. Conversely, the north and the northwestern districts had consistently 5-15ppb

lower NO_2 than the mean. This might be due to the surrounding mountainous areas which constrained the number of roads leading to Seoul. The overall structure of the spatial distribution was similar between interpolation methods (contours and colour codes), however, the spatial patterns varied to some extent. In mid-August, the results of GAM produced a higher and wider concentration in central Seoul than the UK because GAM attempts to find similar values within the error bars (se$\pm$1) whereas UK results tend to print localised patterns. Although the spatial pattern showed some differences, the difference in the actual pollution values was small, meaning that the exposure levels deduced from the two methods might not vary too much.

Compared to the interpolated results, the road overlaid outcomes seem more realistic in terms of weighting pollution to roads and highways. However, a few stations that were closer to roads were overly corrected systematically. Although statistical models cannot take into account atmospheric dispersion in the street canyon scale (Di Sabatino et al., 2008), the modelled outcomes gave improved results relative to the roadside measurements which is promising. For future work, splitting the background-roadside ratio into districts might improve the model's accuracy.

Comparisons Between SI Methods and the Proposed Outcome

Data Density: This study used 57 background stations around Seoul which were unevenly distributed across the study domain (approx. 800km^2). The closest distance was between Jung and Jongno at 1.1km, the furthest between *Daeya* and *Onam* at 44km, and the average of all stations was approximately 29km. Although it is assumed to have a lower accuracy due to wide spatial coverage, there has been little difference between methods in most of the months, meaning that neither constitutes an optimal approach for NO_2 interpolation.

Measuring temporal and spatial variability are useful in areas where long-term temporal data are available for each monitoring station but the spatial distribution is not as dense to cover the flow of air mass (Blanchard et al., 2014). For this study, temporal variability and spatial variability both play a different role but are reciprocal. Temporal variability exhibits the distribution of diurnal and nocturnal pollutants of daily and seasonal change, while spatial variability exhibits the spatial distribution of the mean difference to show how high or low the area is from the mean.

From the NO_2 results, there was a small variation during summer ranging from 15-50ppb with large daily variations partially appearing. The variability in the boxplots can be found in the maps for August and September where the CBD and the southeast of Seoul, known for high traffic volume, are higher than the mean by 1-8ppb. In winter the mean concentration, as well as the temporal variability of NO_2, increased by 38ppb and 85ppb respectively. This increase led to a wider gap between the CBD and southeastern areas and the rest by 5-15ppb (which is why it has a darker colour in winter). Although the temporal variability in summer was small and distributed below the legal limit of 60ppb and the daily cycle showed a continuous pattern of oscillations, the spatial variability supports the boxplot results that higher NO_2 is associated with traffic volume and density (Halonen et al., 2016; Tonne et al., 2008).

PM_{10} had a large temporal and spatial variability in seasonal scale but small variability within a day. A small variation within a day means similarity in the study area. Overall, most days in summer were below the average of 35µg/m^3 with only a few days of instantaneous spikes over 100µg/m^3 in August and September. However, winter PM_{10} levels went up to 60µg/m^3, with the seasonal range of 181µg/m^3. The spatial variability maps revealed that most of the high values were situated in the southwestern area of Seoul and the concentration gradually decreased as it gotets to the northeastern region. The spatial patterns of PM_{10} exhibited a corridor at the high-southwesterly and low-northeasterly trend. It is presumed that the industrial sites in the western and southern parts of Seoul, as well as the transboundary particles, had affected the rise.

Although the results from variability compensated the unknown values to some extent, uncertainty remains. Thus, to get a believable outcome, future study should consider more variables since pollution concentrations are very sensitive in time, location, altitude, and weather, and other anthropogenic impacts. One example is Newcastle, a 'smart city' in the UK, where they have installed detectors inside lampposts in most of the busiest areas of the city that detects air pollution, solar radiation, wind direction, and noise, all stored in the urban observatory for every minute. With this cutting-edge technology and storage, as with Newcastle's case, Seoul could also start building small detectors to supplement the spatial gaps. An alternative way is to use crowdsourced air quality data from social media and mobile applications (Internet of Things, IoT). Sun and Mobasheri (2017) used a mobile app *Strava* that provided air quality news to protect pedestrians' and cyclists' activities and health in

Glasgow. Another study developed a cloud regarding urban big data and health care. The authors proposed a new system, UH-BigDataSys, that interrogates air quality and psychological status from the website and crowdsourced data, then resent useful information such as recommended routes and medical advice (Chen et al., 2018).

Stationarity vs Non-stationarity: UK weakens the stationarity assumption (i.e. second-order stationarity) by allowing the mean values to vary by coordinates while holding the variance constant across the field. However, the author transformed the data for summer, because the NO_2 levels were very low at the period; in other words, the levels were left-skewed and highly peaked on the histogram.

Despite the transformation of the summer data, the outcomes of UK only showed a marginal difference compared to the GAM results. The empirical results contrast with the ideas from Aalto et al. (2013) where the climate interpolation for Finland was more accurately predicted with GAM due to the RMSE. However, it should be considered that RMSE was a monthly factor that only represents a single value for the whole country, but for pollution, it is worthwhile to examine the results at least at the junction level.

Systematic error by locational features: Since SI methods estimate surface values without considering the locational features and environmental factors, it tends to return systematic errors. Based on the stationarity conditions, a semivariogram grouped Jongno-Jung and Gyomun-Jungnang together. The distance between the two locations is, in theory, the same group, however, the NO_2 between Jongno-Jung lies on a busy commercial area, whereas Gyomun-Jungang has a 180m altitude mountain between them (see Figure 3.40).

Execution speed from a local PC: UK takes much more time due to the calculation of semivariograms and data transformation. UK also tends to be a lot slower (i.e. UK: 3-5 mins; GAM 2 mins) in terms of the implementation speed for one period, given the same conditions. The reason might be due to finding the best semivariogram fit, and because it recalculates the value for each pixel location based on the assumption of the second stationarity, although this method is markedly more prone to overfitting. Fluid-dynamic models such as ADMS (Atmospheric Dispersion Modelling System) and CMAQ (The Community Multiscale Air Quality Modeling System), which are frequently used for air pollution modelling, produces a sophisticated map that brings in the complexity of airflow from street canyon layers (Beevers et al., 2013; Hoek et al., 2008; Tonne et al., 2008). The execution speed will take around

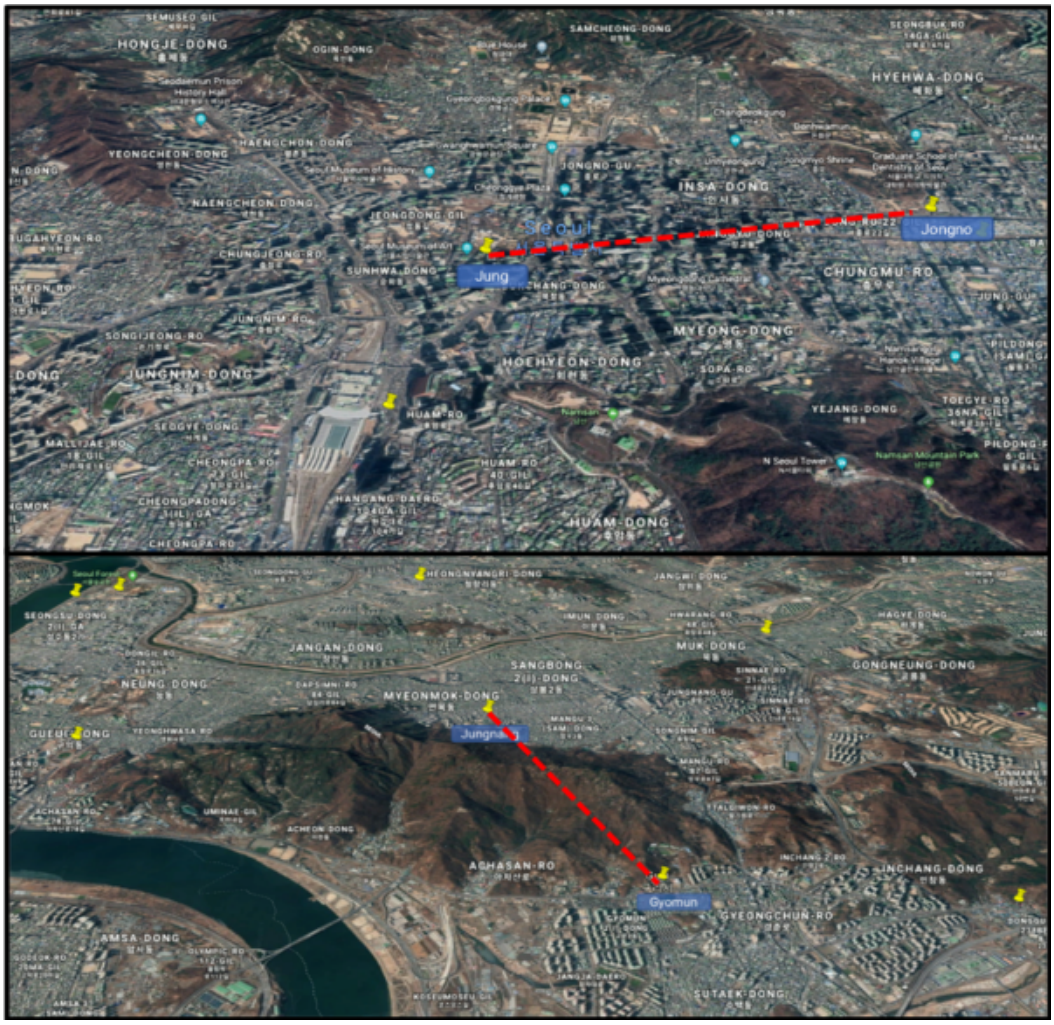


Fig. 3.40 Stations with similar distance but in a different environment

a week to predict one aggregated time step, which is difficult to estimate deleterious effects from daily exposure; however, the emissions, building heights, and the dispersion functions all come together to accurately model the complexity of the airflow, i.e. fluid-dynamic models are better for small scale prediction but takes a long time to produce an outcome. There is a tradeoff between the two approaches. Although the results from this study may not be as accurate as that of dispersion models, this study brings an important contribution in that it requires considerably less computational time to run the model, and the method is open source.

What is a good performing model? Fitting a model using statistical indices - (e.g. R^2, AIC) vs missing the extremes :One of the problems of

interpolating pollution is missing the extremes. This is because the settings of semivariograms cannot capture the small-scale variation that occurs in short spaces, and also the settings of smoothing (wiggliness) function cannot consider every observation value. Section 3.3.3 the chapter was forced to control the nugget to consider the local effects but caused numerous peaks near the stations (a bull's eye effect) but predicted poorly at the unknown areas. This can be a limitation of using a statistical prediction.

Are the existing cross-validation methods worthwhile to compare model performances?: Many published articles have used cross-validation indices such as RMSE, MSE, or MLE to compare model performances (Aalto et al., 2013; Moreno-Jiménez et al., 2016). However, the author discovered that these error indices are strongly associated with the overfitting issues where bull's eye effects of kriging outcomes and the wiggliness of GAM's plate spline makes the map less reliable. In controlling the overfitting issues, this study discovered that the RMSE was lower in most GAM results than that of the UK which follows the result of Aalto et al. (2013), although the difference was small. This might be due to the model's penalised likelihood (the flexibility of wiggliness) that controls the model error, whereas semivariograms tend to use weighted least squares (WLS) or the ordinary least squares (OLS) method to get the prediction. The more important fact here is that the low station density means that there is strong tension between close fits at point locations and the requirement to get a plausible large-scale result.

Summary and future work

Throughout the model, the estimation from both methods can overestimate or underestimate the results since the stations were far away from each other. The results reveal that NO_2 showed a remarkable contrast at the daily road:non-road ratios, whereas PM_{10} had a rather small variance in a daily cycle but had a wide yearly oscillation instead. Our findings show that 12-hour aggregation was temporally appropriate for space-time interpolation and adding road weights on top of the interpolated results was a useful attempt to compare road and background areas.

With regard to the model's performance, neither method was noticeable over one another. The interpolated maps from Figures 3.21-3.24 showed a 10% difference in the edges or the centre but from the boxplots and variability maps, the difference did not exceed 1ppb or 1µg/m^3 on average. While GAM was more flexible, quicker in

delineating the spatial pollutant gradients based on the station locations, and did not require any consideration in anisotropy, UK required substantial time to rigorously adjust to implement semivariograms and justifiable outcomes.

These findings can be further used in modelling population-weighted exposure using different sets of population data. One way is to use mobile population data interrogated from LTE signals (Dewulf et al., 2016; Nyhan et al., 2016), which enables us to see the variability of locations and socioeconomic groups in time and space. Xu et al. (2019) suggested a population-density-weighted exposure because the congestion is what generates more pollution and health risks. They used a new method that combined spatial interpolation on road segments, OD surveys of travel modes and destinations, and a traffic assignment model that estimates the level of traffic travel time.

In subsequent chapters, the studies will make changes to two objects. First, using the advantage of simulation, the studies will consider small-scale variation of pollution patterns that will dynamically change time and space. Moreover, while the interpolation removed the peak values to obtain a spatial structure, the simulation models will make use of the extreme data that was actually seen from the station data. Another improvement is to add activity patterns of individuals that will increase the accuracy of the location where the individual was exposed.

Chapter 4

An Agent-Based Assessment of Health Vulnerability to Long-Term Particulate Exposure in Seoul Districts

Chapter Overview

This chapter evaluates the health effects of long-term exposure to PM_{10} in Seoul districts. An in-silico agent-based model (ABM) is used to simulate travel patterns of individuals according to their origins and destinations. During the simulation, each person, with their inherent socio-economic attributes and allocated origin and destination location, is assumed to commute to and from the same places for 10 consecutive years while a nominal measure of their health is set to decrease whenever the concentration of PM_{10} exceeds the national standard. This chapter differentiates movement between demographic groups and the range of recovery depending on a proxy of socioeconomic status, i.e. housing values. Three scenarios of seasonal PM_{10} change (business as usual: BAU, exponential increase: INC, and exponential decrease: DEC) and three scenarios of resilience were investigated, comparing the vulnerability rate both between and within each district. The first result shows that the vulnerability of groups across Seoul districts, including those aged over 65, aged under 15, and with a low education level, increased sharply after 7,000 ticks (with each tick corresponding

to 1 day) and road proximity caused an additional health decline. This implies that disparities in health outcomes can be associated with demographic groups, especially when the group is exposed over a long period. The second finding from the local-scale spatial analysis of two contrasting districts indicates that most sub-districts in the wealthy Gangnam area showed more variation both spatially and in different resilience scenarios, whereas the less well-off Gwanak areas showed a uniform pattern regardless of earlier prevention. The implication for policy is that, while some areas, such as Gwanak, clearly require urgent mitigating action, areas like Gangnam may show a greater response to simpler corrections but aggregating up to the district scale may miss particular areas that are more at risk. Future work should consider other pollutants as well as more sophisticated population and pollution modelling, coupled with explicit representation of transport and more careful treatment of individual doses and the associated health responses.

Keywords: PM_{10}, Exposure, Health vulnerability, Agent-based model (ABM), Seoul

Note: Published Work

This chapter has been published in the Journal of Artificial Society and Social Simulation (JASSS) entitled, An Agent-Based Assessment of Health Vulnerability to Long-Term Particulate Exposure in Seoul Districts (2019) (https://doi.org/10.18564/jasss.3940). Substantive areas of modelling and writing has been done by myself, and the co-author (PhD supervisor) has commented on and edited the text and provided some conceptual input. Compared to the journal article, this chapter made some changes in the movement patterns and expanded the study area from two sample districts to 25 districts.

4.1 Introduction

In recent decades, urbanisation has consistently influenced economic benefits which resulted in population growth and the increased demands on resources. However, this figure has triggered heavy traffic congestion and acute air quality (Guarnieri and Balmes, 2014). Poor ambient and indoor air quality may result in more illnesses and deaths.

The resulting health issues may vary according to ambient pollution levels but also by people's socioeconomic status. Current studies have discovered that the risk to health from poor air quality is associated with vulnerable groups, particularly those who are young and old, socially deprived, and live near polluted areas (David and Don, 2012; Kan and Chen, 2004; Kan et al., 2008; O'Neill et al., 2003). However, the methods that they have used only considered annually averaged pollution levels, or were temporally too short to discover the effects between the demographic groups.

Another concern is how individual behaviour and socioeconomic status can affect and be affected by pollution exposure. Many public health studies have investigated personal exposure patterns using spatial interpolations (in Chapter 3), surveys, mobile gadgets, or backpack sensors, but less attention has been paid to determining how different social groups move in a continuous space or which trajectories through which particular areas cause high exposure. Without considering human behaviours, there would be less accuracy reflected in terms of air quality where people are breathing. These problems, therefore, have highlighted the need to discover vulnerabilities to pollution across populations and to develop tools to help prevent future health problems.

Quantifying health response to pollution exposure already involves a high degree of uncertainty (Dias and Tchepel, 2018), and studying it at the population level introduces further assumptions in terms of social and individual characteristics and the structural dynamics of urban growth. However, although the impacts on health may be hard to deal with, the patterns of exposure may become clearer when modelled at the individual level, and it seems plausible that higher exposure levels are more likely to lead to disease.

Given that agent-based models (ABMs) are capable of simulating travel patterns with some structure, this chapter **investigates the potential health vulnerability to long-term particulate exposure in Seoul**. The specific research questions are:

- How does the socioeconomic background potentially affect health outcomes?
- How could these health levels change under different pollution development scenarios?

To answer these questions, we developed an *in-silico* agent-based model (ABM) that measures population vulnerability through various scenarios in all 25 districts of Seoul.

The remainder of this study is structured as follows. Section 4.2 briefly describes the background of individual exposure studies and reviews ABMs that are closely associated with health. Section 4.3 provides an introduction to our study sites in Seoul, along with a list of our datasets. In Section 4.4, the setup processes and the scenarios of the agent-based model are documented. Section 4.5 examines the sensitivity of the calibrated parameters that are key for this chapter. Section 4.6 compares a measure of the potential risk to health by district, sub-district, age, and education between the 25 districts and presents a map of their spatial distribution. Concluding remarks and suggestions for future work are then discussed in Section 4.7.

4.2 Related Works

4.2.1 The Spatial and Temporal Dynamics of Air Pollution Exposure

Exposure to air pollution takes place when a person inhales harmful pollutants in a location where there are perceptible amounts of toxins in the air. According to Ott et al. (2006), there are two fundamental points to check to model the level of exposure to air pollution effectively: 1) the location of people who are being exposed, and 2) the heterogeneity of pollution concentration in different locations. These questions led to a simple formula, `exposure = concentration * time`, which in turn was developed into advanced models considering daily behaviour patterns or multiple-station data (Dias and Tchepel, 2018).

Previous studies related to quantifying exposure and health outcomes can be classified into two main streams. Population exposure studies have attempted to associate air pollution with socioeconomic factors, presuming that social status associated with pollution exposure might have resulted in increased mortality (or hospital admissions) (Chen et al., 2007; Halonen et al., 2016; Wang et al., 2008). These studies used statistical models, including Poisson regression models and generalised linear models, to measure the exposure using patient data or census data. However, using a daily aggregated level of exposure possibly smoothed out the immediate effects of large, short-term fluctuations in pollution levels, as well as the characteristic effects of an individual in particular locations, which in turn biased the results.

To avoid this ecological fallacy (i.e. assuming individual behaviour can be deduced from data about groups) recent studies have measured personal exposure by exploiting backpack sensors (Langrish et al., 2009; Steinle et al., 2013), surveys (Beevers et al., 2013; Van Ryswyk et al., 2013), or mobile phones that combine with biomarkers, to demonstrate that where people go is important for their pollution exposure (Dewulf et al., 2016; Dias and Tchepel, 2018; Nyhan et al., 2016).

Although this approach has widened our spectrum of exposure assessment, many questions remain to be answered. Firstly, without having personal information, it is difficult to know whether people with different personal characteristics (e.g. age, sex, and genetic make-up) or social classes (such as wealth or education level) behave differently, use different modes of transport, or move close to more polluted areas, for example, multilane roads and junctions. Additionally, these studies have often been severely time-limited. For example, Dewulf et al. (2016) implemented two days of mobility data available for research, and similarly Nyhan et al. (2016) measured a week of population-wide $PM_{2.5}$ exposure. In cases of personal exposure assessment studies in which the causes of symptoms are non-communicable, disease may develop over many years and the exposure levels are subject to the high variability of space and time, studies need to be long-term, allowing a history of pollution contact to be developed.

4.2.2 Agent-Based Models (ABMs) in Relation to Pollution and Health

Agent-based models (ABMs), particularly in pollution studies, have strengths in exploring finely resolved dynamics of long-term individual trajectories in both space and time and the ability to link these directly to the generation of the pollution field, particularly the very significant part of this that relates to transport. This means that we can not only examine passive exposure to pollution when moving to destinations but also incorporate adaptive behavioural changes and consider possible associated consequences. For example, awareness of exposure to high pollution levels may lead drivers to attempt to alter their commute, thereby changing the places and times of pollution generation and feeding back into further exposure.

Furthermore, we can envisage policy experiments to anticipate the possible effects of road closures, traffic restrictions, or road improvement schemes on the generation of pollution outside protected zones and thereby help to prevent unintended consequences

of attempts to control pollution in highly affected areas. The effects of changing transport mode can also be considered; for example, car drivers are typically thought to be more highly exposed to pollution than pedestrians in the same area (Vreeland et al., 2017).

Finally, we can directly address matters of social justice – those who are disadvantaged typically have less power to effect change for themselves than those who are more well-off and are often less highly exposed to environmental degradation. With an ABM, we can conduct scenario experiments aimed at improving conditions at the individual level, taking into account age, education level, living and work spaces, and the ability to travel, including the adaptive behaviour of the agents in response to policy changes. In all, the flexibility of ABMs and the ability to estimate uncertainty in scenarios should help to guide our understanding of how pollutant exposure might translate into disease.

Although previous models have thoroughly investigated the detrimental pollution impacts on human health, relatively few have used agent-based modelling (ABM). ABM has perhaps been underused in environmental health disciplines due to either the complexity of dealing with heterogeneous individuals moving through complicated geographies or the lack of sufficiently spatially resolved micro-data.

One simple air pollution exposure ABM was a prototype model called Urban Suite-Pollution (Felsen and Wilensky, 2007). The idea was to examine the competition between predators (fixed locations emitting pollution) and prey (people's reproduction) in an enclosed landscape. This model can lead to a form in which the healthy population is sharply reduced due to the pollution and ageing effects. Agents can mitigate the pollution effects by planting trees, which are envisaged as being able to help clean up the pollution diffused from point-source power plants. However, after testing the model's sensitivity, it became clear that fluctuations in the population can be dominated as much by the random walk of the population size, when the birth and death rates are nearly equal, as by the effects of pollution. Moreover, there is no link between the population size or agent locations and the generation of pollution, and the effects of other built infrastructure are lacking (in particular roads). Finally, the pollution spreads purely by diffusion from each point; in practice, the effects of wind, rainout, and urban street canyons are likely to be significant. More realistic geographies are needed to make further progress.

More recently, David and Don (2012) developed an urban pollution model (EPICast-API) that calculates human movement, time usage, dose, and response to estimate the mortality rate, respiratory hospital admissions, and emergency room visits in the greater metropolitan area of Sydney, Australia. As far as we are aware, this was the first pragmatic ABM to use an entire urban population to investigate pollution vulnerability. However, exploring the relationship between mortality statistics at a coarse grid scale ($1.5km^2$) and only for a single year may miss the effects of spatial variability and cumulative exposure. In addition, their study did not decompose the results by social class.

While the complexities of the evolution of the urban air pollution field combined with modelling people with different backgrounds present a strong challenge for studies of this type, the small number of existing pollution simulations with ABM suggest that there is hope that more detailed simulations will help us to make progress. Here we attempt to improve on the randomness of agents' behaviour from the Urban Suite-Pollution model by using origin-destination data from travel surveys, use measured data to estimate the pollution field and increase the spatial resolution of the pollution grid and the temporal resolution relative to the *EPICast-API model*, whilst also breaking down the population by personal and social characteristics.

4.3 Data Collection

Table 4.1 provides an overview of the variables used in this initial model. Station data, individual personal attributes, road layouts, land prices, and an origin-destination matrix for daily travel was collected.

Table 4.1 List of Variables

Components	Variable	Value	References
Pollution	PM_{10}	Numeric	National Institute of Environmental Research
Sensitivity	Age	Numeric	Korean Statistical Integrated System (KOSIS)
	Education	Dichotomous	Korean Statistical Integrated System (KOSIS)
	Road proximity	Dichotomous	Korean Transport Database (KTDB)
	Official land price	Grade	National Spatial Data Integration (NSDI)
Movement	Origin-destination	Numeric	Korea Transport Database (KTDB)
	Land cover	Categorical	National Institute of Environmental Research

4.3.1 Pollutants

PM_{10} was selected not only because the data was available but also because of the growing emphasis of PM_{10} impacts as a first-class carcinogen that can cause pulmonary disease, cardiovascular problems, and stroke (IARC, 2013; Kan et al., 2008; Lee et al., 2014; Loomis et al., 2013). Hourly PM_{10} data were collected from the nearest background stations from each district between 1 January 2010 and 31 December 2015 then grouped into home hours (assumed to be 20:00–08:00) and working hours (09:00–19:00), as illustrated in Figure 4.1. To account for some gaps in the time series (e.g. 2.15% in Gangnam and 785 hours 1.49% in Gwanak), this study inferred the missing values using a Kalman algorithm for each season from an `ImputeTS` package in R (Moritz and Bartz-Beielstein, 2017).

At this stage of the model development, agents' movement was not tracked alongside the road system. However, this study included a measure of road proximity as a means of demonstrating the negative impact of enhanced emissions near roadsides, irrespective of social categories. The data were retrieved from the Intelligent Transport Systems (ITS, http://its.go.kr/) institution run by the Korean Ministry of Land and Transport.

4.3.2 Demographics

For this research, age and education were the characteristics that featured in pollution sensitivity. Age was included, as health risks to air pollution can depend on personal physical condition, and in general are prominent in vulnerable age groups (Pearce et al., 2006). Population data were retrieved from the 2010 Census of the Korean Statistical Office (http://kosis.kr/), which were grouped into five-year intervals (e.g. 0-4, 5-9, 10-14, ..., 80-84, over 85). The age intervals were kept as it was, but ages under 6 were removed in order to align with the educational population that was surveyed from the age of 6. Thus, 17 groups were used in this study (see Figure 4.2).

The education level was included to represent the awareness of potential harm from pollution. (see Figure 4.3) The 2010 census surveyed education status above 6 years of age by district and provided data in 24 categories: No Education, Primary, Middle, High, College, University, Masters, and PhD – including dropouts, current students, and graduates. For simplicity, they were aggregated into eight status groups, and the district level populations were converted into fractions to approximate the census.

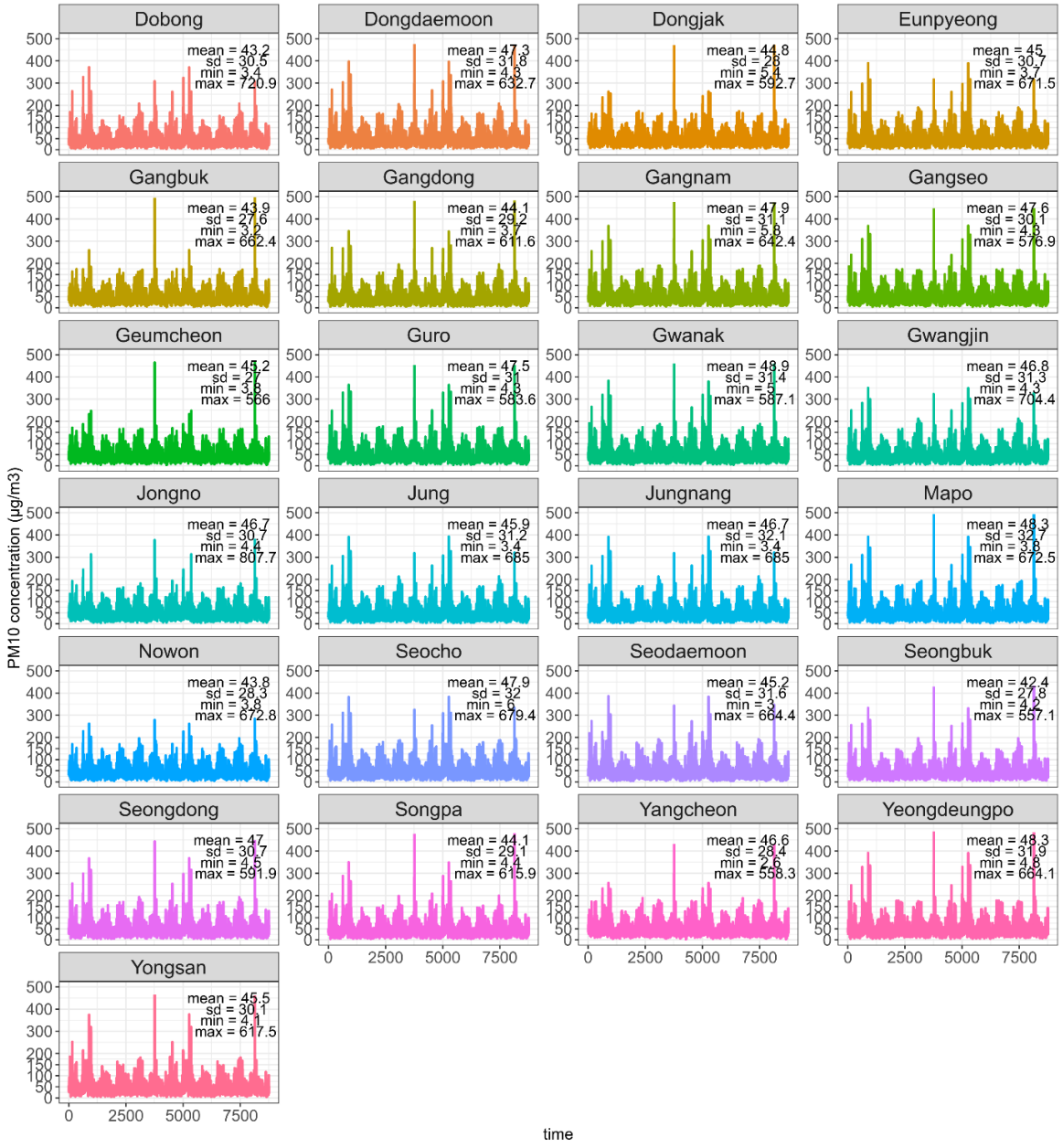


Fig. 4.1 Time-series of PM_{10} changes in 25 Seoul districts in 2010-2015. One unit of the x-axis is 12 hours, and the final date, December 31st 2015 would be the 8764th tick. While the peaks vary by districts and years, the majority of the monitoring results remain below 150µg/m^3

The 2010 census surveyed education status above 6 years of age by districts, and provided data in 24 categories - No education, Primary, Middle, High, College, University, Masters, and PhD – including dropouts, current students, and graduates

(see Figure 4.3). For simplicity, we aggregated into eight groups of their status at which the individual was surveyed, and we converted the district level population into fractions to approximate the census.

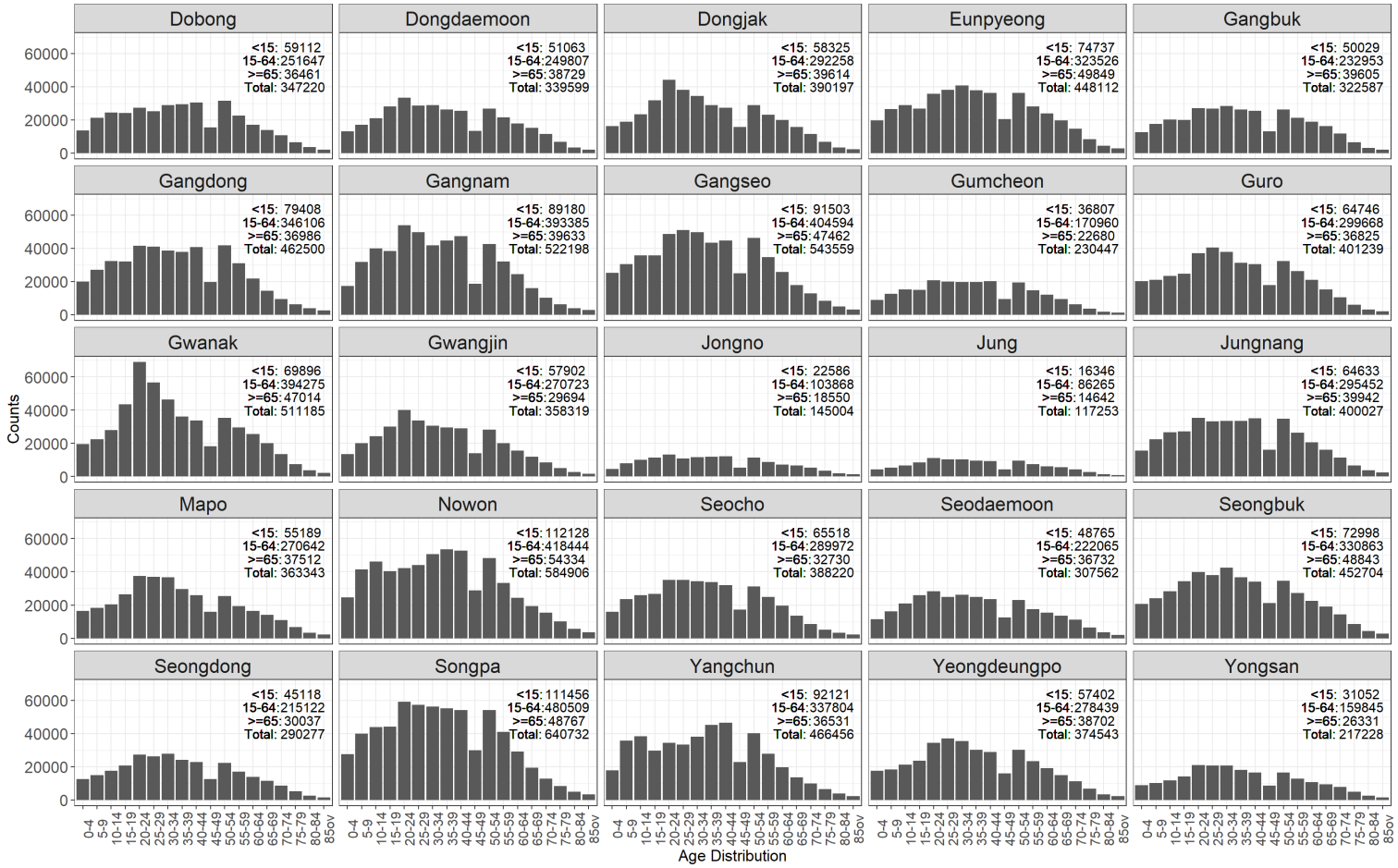


Fig. 4.2 Age distribution for each Seoul district in 2010. Apart from the CBD districts, Jung and Jongno, most of the districts have population distributed between 300,000 and 640,000. Except a dip portrayed in the 40-44 age group in all districts, the population naturally showed declining trend from the mid-40s.

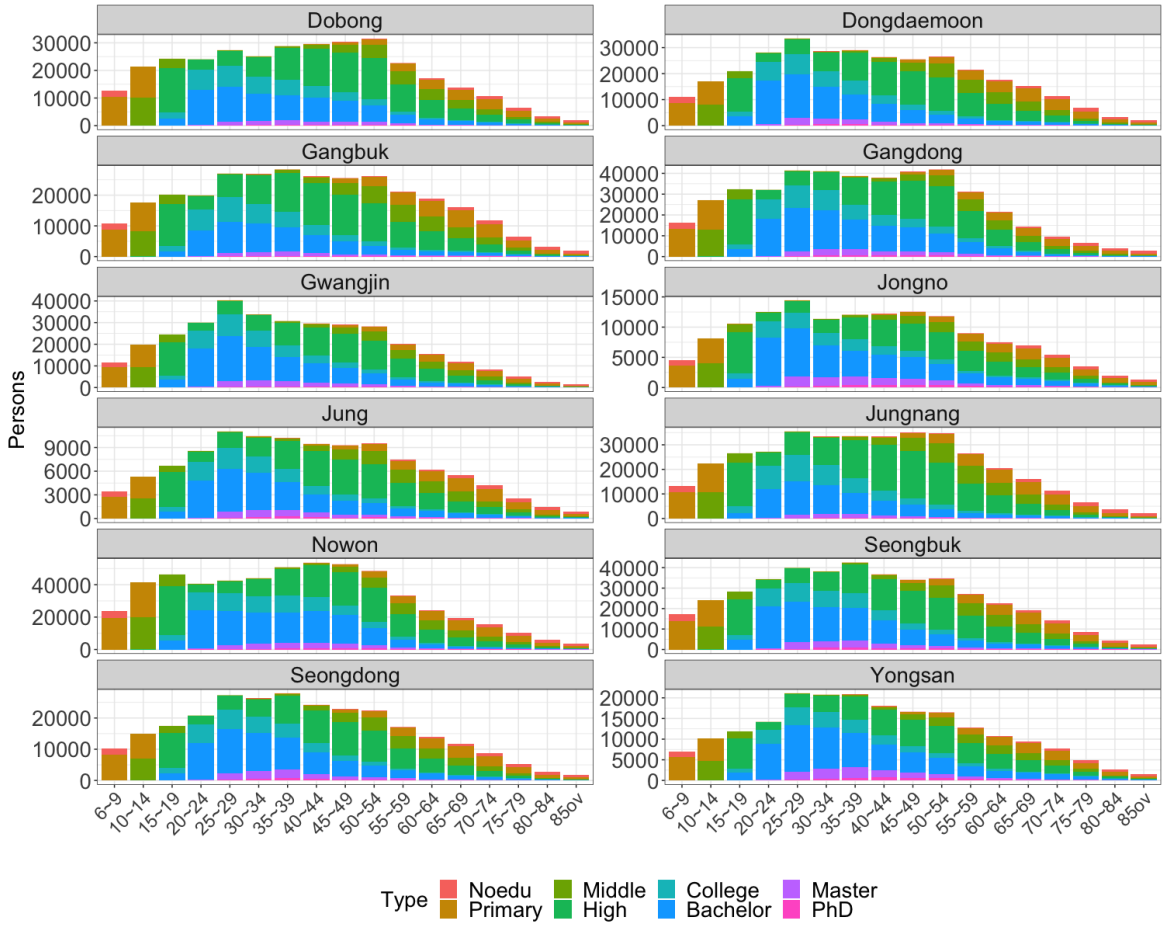


Fig. 4.3 Education levels by age distribution in 2010 across 12 districts: Jongno, Jung, Yongsan, Seongdong, Gwangjin, Dongdaemoon, Jungnang, Seongbuk, Gangdong, Gangbuk, Dobong, and Nowon. The population of each district is nearly identical to the age plot illustrated in Figure 4.2 because it excludes age group below 5 years of age or those whose education was unreported. The highest education group across Seoul is high school, college and university(bachelor) graduates.

4.3.3 Health Recovery

Official Land Price was selected as a proxy for the rate of recovery from pollution impacts. Coffee et al. (2013) attempted to show a relationship between residential properties with high socioeconomic status and health as part of a study of housing properties and social well-being. The reason for using residential property was not only because it represents immovable and location-specific capital but also because of its expression of price in the economy: those who are more able to afford higher-priced

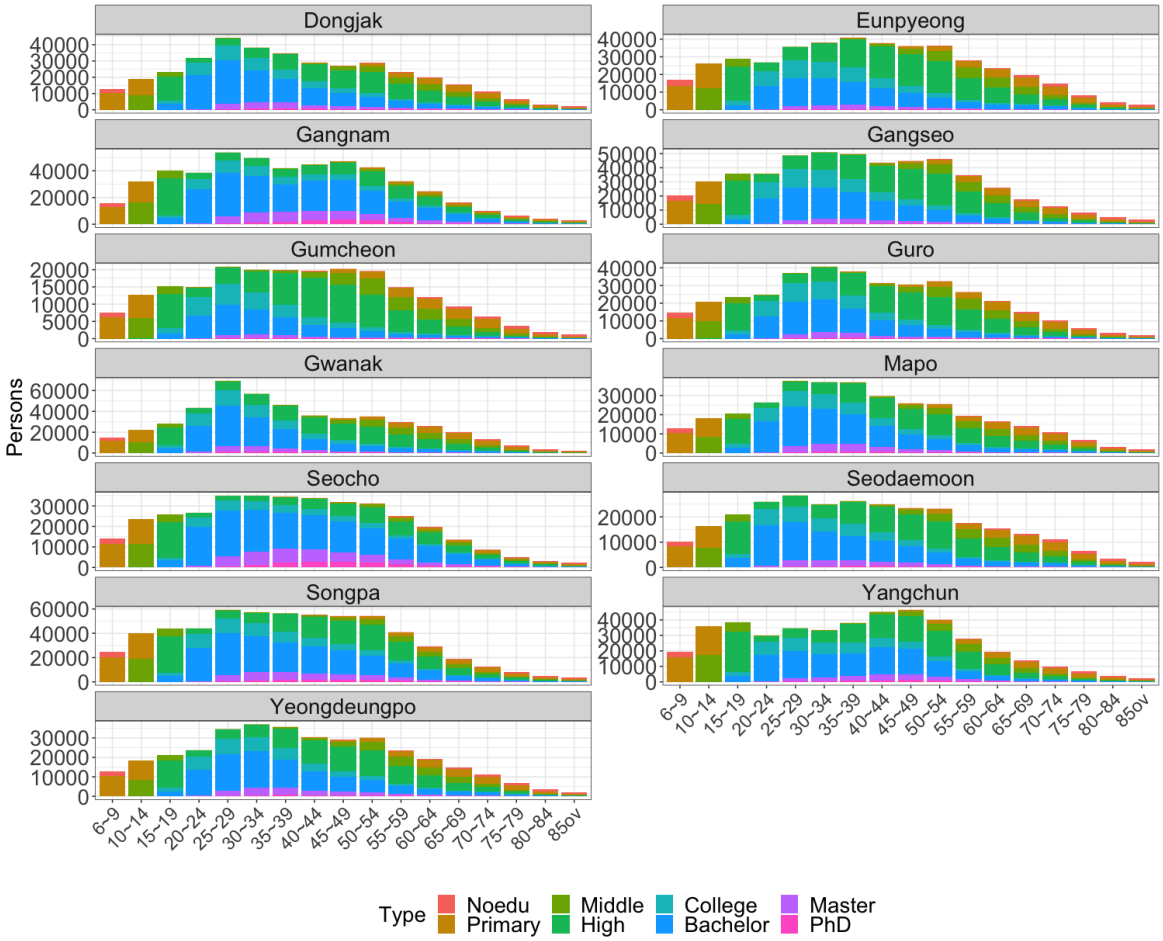
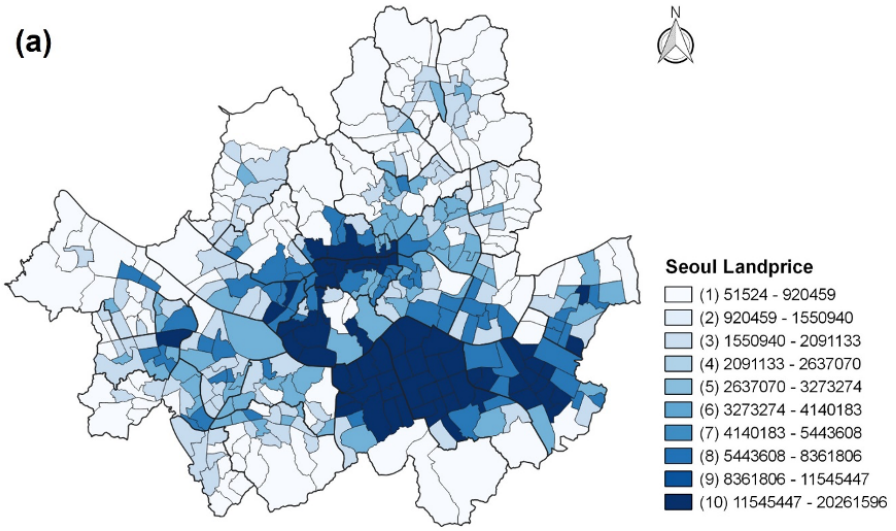


Fig. 4.4 Education levels by age distribution in 2010 across 13 districts: Eunpyeong, Seodaemoon, Mapo, Yangchun, Gangseo, Guro, Gumcheon, Yeongdeungpo, Dongjak, Gwanak, Seocho, Gangnam, Songpa. A contiunous plot from Figure 4.3

housing are also perhaps better able to access health care or to adapt their lifestyle to compensate for high pollution levels.

The 2015 Official Land Value data were retrieved from the National Spatial Data Integration website (NSDI, http://www.nsdi.go.kr/). The data were given in tables of prices by census block units. Using the natural break method, the block units were aggregated to the sub-district level, and then the price distribution was categorised into 10 sub-divisions (see Figure 4.5(a)). Figure 4.5(b) indicates the hypothetical health changes determined by the land prices. It is assumed that agents' nominal health level can recover up to a maximum adaptive capacity level at a rate that is higher for those living in areas with higher property values.

(a)



(b)

Class	Land Price (in Korean Won)	Health
1	< 920,459	+ 0.01
2	920,459 - 1,550,940	+ 0.01
3	1,550,940 - 2,091,133	+ 0.02
4	2,091,133 - 2,637,070	+ 0.04
5	2,637,070 - 3,273,274	+ 0.05
6	3,273,274 - 4,140,183	+ 0.05
7	4,140,183 - 5,443,608	+ 0.06
8	5,443,608 - 8,361,806	+ 0.07
9	8,361,806 - 11,545,447	+ 0.12
10	11,545,447 - 20,261,596	+ 0.15

Fig. 4.5 Official land prices of Seoul sub-districts in 2015 in sub-district scale (a), and table of hypothetical health changes determined by land price in Korean currency (£1≈₩1,500) (b)

4.3.4 Movement

Origin-destination matrices for all districts were downloaded from the Korean Transportation Database (https://www.ktdb.go.kr) and converted into fractions, so that home locations and associated work locations could be allocated to a sub-sample of the total population.

Since OD matrices produce a coarse temporal resolution, namely two locations per day, the study assumed the population moves across landcover of either Residential

(110), Commercial (120), or Traffic (roadside) (150) areas that would be likely for home and work locations (see Figure 4.6).

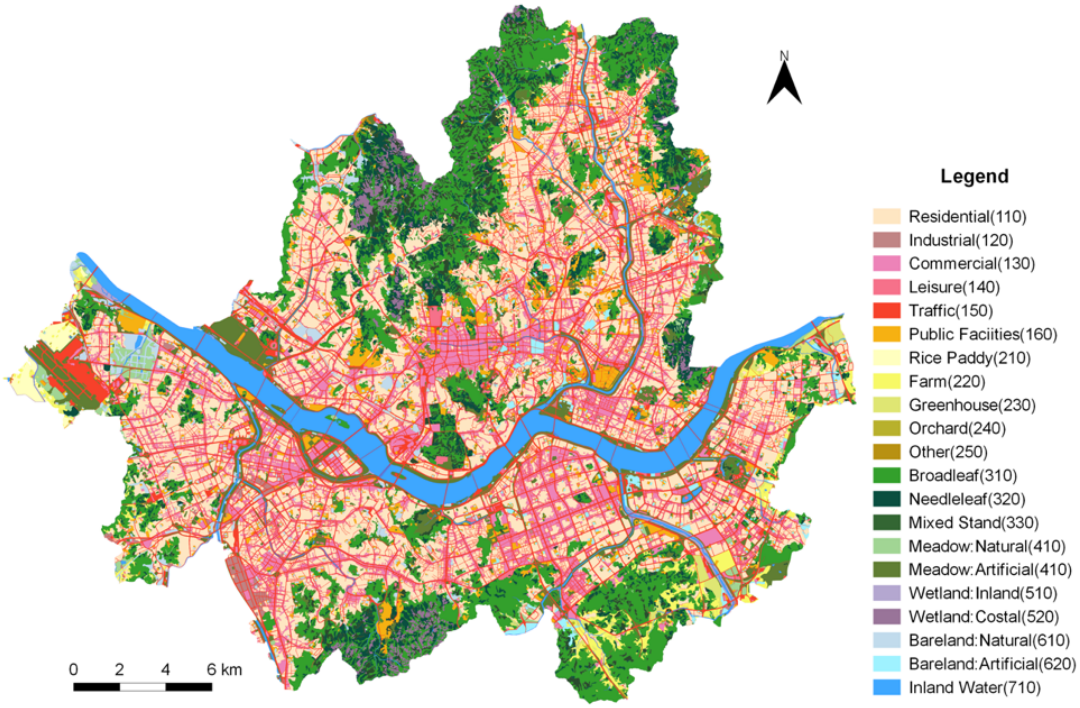


Fig. 4.6 The landcover of Seoul (2014) provided by the National Institute of Environmental Research. This research exported Residential (110), Commercial (120), and Traffic (150) for the agent's movement (colourcoded by the author)

4.4 Model Description

In this section, we describe our pollution exposure model based on the ODD protocol (Grimm et al., 2010). The source codes, along with the extended ODD protocol, are available from CoMSES[1]. The model was implemented in `NetLogo 6.0.4`, with each district runs as a separate self-contained model: this means that some agents with home locations within a district but work destinations in another district had to be excluded from consideration. We implemented nine scenarios for 20 iterations each

[1]https://www.comses.net/codebases/cb6c2243-fb44-4543-a372-6fee5f034c40/releases/1.1.0/

in `Behaviorspace` embedded in `NetLogo`, and then the study analysed the outcomes with `tidyverse` in `R 4.0.2` (Wickham, 2019).

4.4.1 Model Purpose

The model's objective is to understand the cumulative effects on population vulnerability from exposure to PM_{10} by different age and educational groups in Seoul districts.

4.4.2 Entities, State Variables, and Scales

For the sake of speed, this study used a 5% population sample of agents (individuals) to generate a simple synthetic population in Seoul districts. For example, the sampled population of Gwanak consisted of 491,554 persons, of whom 50,265 (10.2%) were aged under 15 (excluding infants under 5), 394,275 (80.2%) were aged between 15 and 64, and 47,014 (9.6%) were aged 65 or above. In the model, each group was converted into 2,427, 18,775, and 2,250 people respectively and in total of 23,452 agents.

Each agent has a list of attributes that includes the name of the home location, coordinates of the home location, name of the destination, coordinates of the destination, age group (i.e. young, active, or old), education level, and health. Every agent will move to his or her given destination location during working hours and return home again on alternate model ticks (see more details in Section 4.4.3).

It was assumed that the agents have not experienced any previous exposure, as this study lacks either health records or exposure histories for the population. Our simulation is thus only able to track the likely rate at which exposure effects accumulate over time in a given district relative to others. Depending on the way in which this simulated history works, this chapter hopes to be able to estimate the accumulated risk that an unexposed population would have developed over the period of the simulation. Each agent is assigned a nominal 'health' level, which is an integer with an initial value of 300. Depending on their socioeconomic status, they will lose health when exposed to a PM_{10} over a threshold near $100\mu g/m^3$. This study chose this as it is a Korean hourly ambient air quality standard, although the lack of data on how this relates to disease means that the relation to the actual health impact is not entirely clear. This study assumed, consistent with the idea of an air quality standard, that the adverse

effects on health only begin to operate when pollution is above this threshold. More details on this matter are given below.

In addition to PM_{10} exposure, the demographic sensitivity was used as a modifier of the adverse factors degrading the agents' health and used the land price as a recovery factor. Sensitivity is composed of two socioeconomic attributes of an agent that control the degree of health loss. One determinant is age, which works as a proxy for physical health variance, assuming that very young or very old population members are more likely to suffer when exposed to pollutants. The other is educational level, designed to take into account a possible lack of pollution awareness amongst those with lower educational depth. By contrast, the land price is a proxy for health recovery based on the idea that medical facilities (e.g. hospitals, clinics, and pharmacies) have greater chances of being located in areas with higher property values or that higher values may mean a greater ability to avoid polluted air, as mentioned above. The model does not consider any transmission effects between individuals or environments, because the diseases resulting from pollution exposure are non-communicable, although, in practice, compromised pulmonary systems may make agents with high exposure more likely to suffer ill-effects from viral or bacterial diseases.

4.4.3 Process Overview and Scheduling

Figure 4.7 shows the conceptual outline of the model. During the set-up process, every agent is assigned a fixed home name (sub-district) and home patch as well as a destination name and patch. The destination names and patches differ by age group. Agents aged between 15 and 65, also known as the economically active population, will move to their destination patches according to the fraction moving to a given destination in the origin-destination (OD) matrix. Those who commute to other districts are allocated to dummy patches outside the district during working hours – these agents are not included in the overall statistics, as we do not have data for them during the day.

Agents aged under 15 will move to a random patch within radius 3, while those aged over 65 will move to a random patch within radius 1: this is intended to represent a more restricted range of movement for this fraction of the population. As we were not modelling the traffic flow, we simplified movement by translating the agents to their destination patch during the day (one tick) and back to the home patch at night

(the next tick). Every agent will start without any medical history and with full health; all are susceptible to local levels of PM_{10} within their current patch. An additional weight for PM_{10} is added if the agents are near a road, assuming that the particles from tailpipe exhausts or tyre dust are generated near roads (see the details in section 4.4.5). An agent will lose health when exposed to a patch exceeding 100µg/m^3 of PM_{10}. Agents in the youngest or oldest age groups and those with lower educational levels receive an additional penalty. Recovery can occur according to local property values up to a maximum adaptive capacity (see the details in section 4.4.6).

As the agents continue to lose health from a high PM_{10}, we assumed that their health status can change suddenly when they cross a given threshold. Those with health less than two-thirds of the starting value change status to 'highly exposed', and those with health less than one-third of the initial value are labelled 'at risk'. The model terminates if either all the agents turn to the 'at risk' status or the end of year 12 is reached, that is, tick 8764.

4.4.4 Pedestrians' Decision-Making Process

The agents in the model have little by way of reasoning ability. Their behaviour is entirely driven by a simple schedule. Agents, in common, follow the hypotheses mentioned below:

- An agent's birth, death, and ageing are not considered
- Agents have minimal cognitive representation but understand their area boundaries
- 1 tick is equivalent to half a day, that is, working and home hours
- Every agent starts with a health status of 300, but this drops when the agent is exposed to pollution
- Agents move from and to either residential, commercial, and traffic areas
- Agents commute to the same location until the simulation ends
- For visualisation purposes, if the health status of an agent drops below 200, the agent's colour turns purple; when it drops below 100, the colour turns red
- If an agent's health reaches 0, he or she will be sent to the hospital for treatment

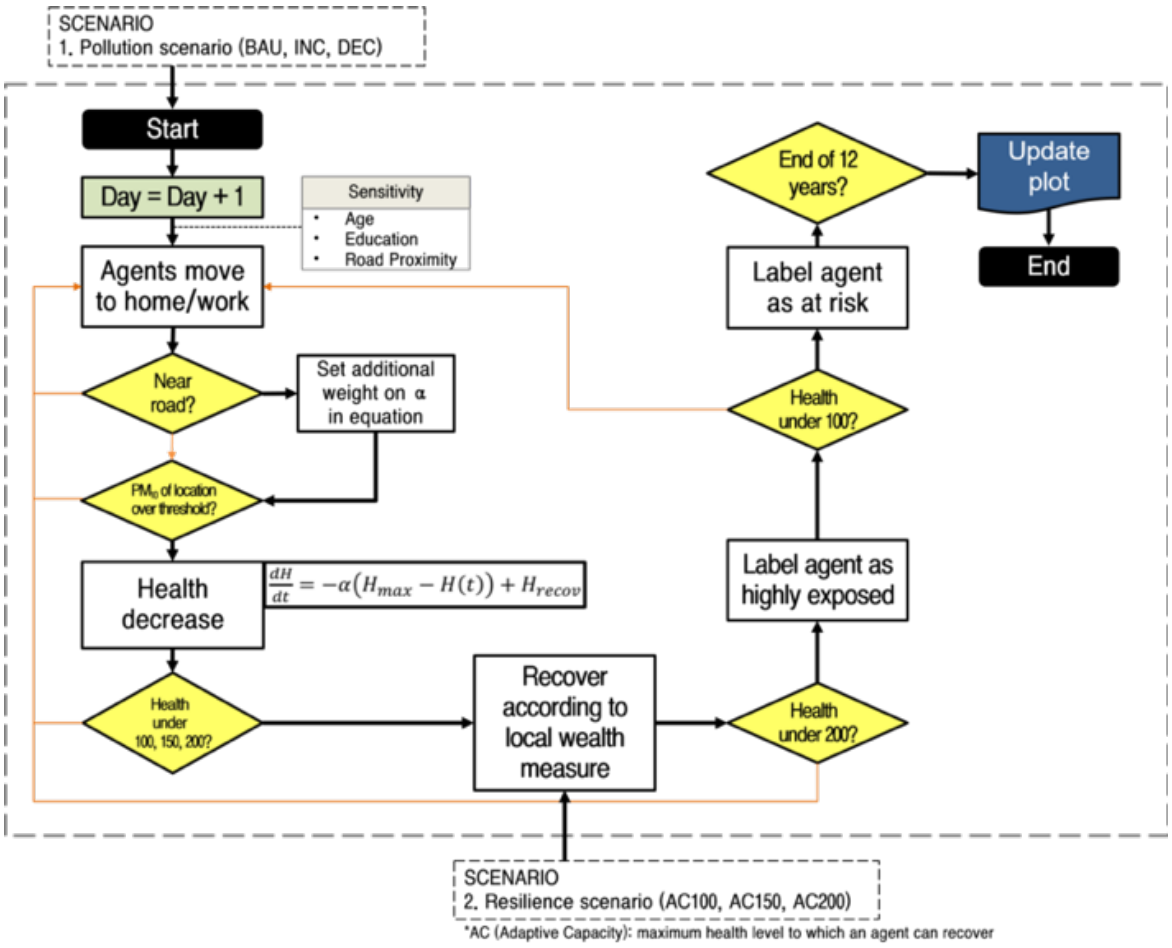


Fig. 4.7 The implementation algorithm for each period

- All agents stop if the system reaches 8764 ticks (equivalent to 12 years) or if the 'at risk' population reaches 100% of the total population

Given the lack of address information, every agent was randomly distributed in their home location based on the census. Agents aged 15–64 commute within a sub-district but can also move to different sub-districts derived from the origin-destination matrix. Movement is random within these constraints but with an additional tendency to select destination patches of commercial areas or near roads within permitted sub-districts randomly as a way of approximating transportation in the absence of a full traffic model. By contrast, the movement range of the young and the elderly is restricted close to their origin. This is based on the idea that most children go to schools and participate in after-school activities near home and that the elderly are assumed to travel to distant places less frequently. This study did not consider agents who commute

to different districts outside the study area or the mode of transport for trips (so there is no accounting for differences between pedestrians, car travel, or buses or for the fact that the pollution levels may be lower within buildings). The effect of roads proves to be significant in determining the relative outcomes between groups.

4.4.5 Atmospheric settings

Most pollutants vary quite strongly across location and time. Given each district had only one background station, estimations of personal exposure based on uniform air quality might overlook a significant amount of variance of pollution across space and time and can under- or overestimate personal exposure levels, particularly when the study region is larger than a census block level (Dias and Tchepel, 2018). Hence, rather than allocating a uniform value homogeneously to a region, we allowed each patch to select one value randomly from the daily pollution field measured at the local background station (see Figure 4.8). For example, on the first day of January 2010, each patch would randomly select one of the PM_{10} values between 09:00 and 19:00. This allows for the likely spatial variability between patches to be preserved rather than all the patches varying in exactly the same way as the local station.

This chapter also applied an additional weight factor of pollution to road exposure, following the evidence that adverse effects on health are more likely to occur in proximity to roads. One of the pieces of evidence from Harrison et al. (2004) involves an average mass increment of 11.5µg/m^3 of PM_{10} and an 8.0µg/m^3 increment of $PM_{2.5}$ between roadside (daily mean 34.7µg/m^3 PM_{10}) and urban background areas (daily mean 23.2µg/m^3) measured at three sample sites in London and one in Leeds. Higher chances of acute illness can happen because of the direct inhalation of toxic fumes inside a vehicle or walking close to vehicle exhaust. It was discovered that the urban roadside stations in the districts between 2010 and 2015 showed a 40% and 38% higher concentration of PM_{10} than the background stations. Thus, we applied these levels to the road patches in each district.

4.4.6 Measurement of Exposure

In accordance with the PM_{10} concentration criteria provided by the Korean hourly ambient air quality standard, the air quality is considered to be harmful when it exceeds

(a)

Randomly choose one of the values in the first row

No	Date	Work	h1	h2	h3	h4	h5	h6	h7	h8	h9	h10	h11	h12	h13
1	2010-01-01	home	30	35	36	31	33	36	28	38	38	52	43	43	39
2	2010-01-01	work	37	33	31	41	29	41	36	36	37	35	37		
3	2010-01-02	home	39	37	38	40	43	39	41	40	99	85	83	77	70
4	2010-01-02	work	46	44	47	42	39	42	47	55	68	75	102		
5	2010-01-03	home	64	62	59	56	54	52	47	40	46	59	61	64	65
6	2010-01-03	work	49	36	33	41	45	39	46	37	44	50	43		
7	2010-01-04	home	57	48	54	51	44	37	35	36	34	62	54	44	45
8	2010-01-04	work	25	26	28	24	25	27	31	30	33	27	26		
9	2010-01-05	home	46	56	53	60	61	62	62	59	40	33	31	31	37
10	2010-01-05	work	65	63	56	58	41	46	68	64	47	29	41		

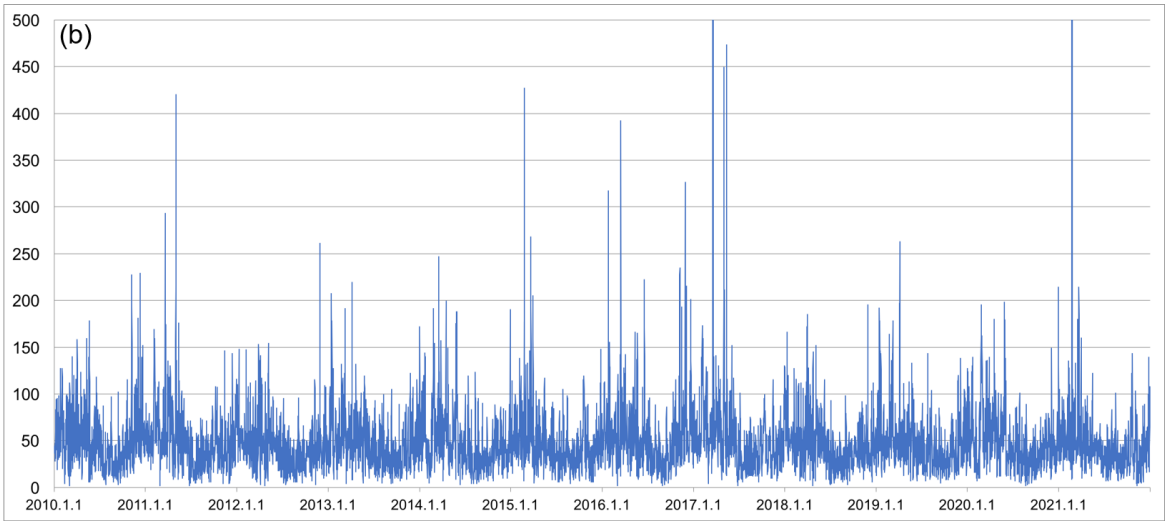


Fig. 4.8 (a) The stochastic process for selecting PM_{10} for each patch per row and (b) a simulated PM_{10} result plot from a random location of a BAU simulation

100µg/m^3, although the lack of data on how this is linked to disease means that the relation to the actual health impact is not entirely clear.

As mentioned above, since the diseases resulting from pollution exposure are generally non-communicable, at least to the first order, the model does not consider any transmission effects between individuals but considers the effects resulting from continuous interactions between individual spatial trajectories constrained by daily activity patterns, with simulated spatial micro-scale atmospheric pollution distributions.

It was assumed that, compared with a healthy person, a person suffering from disease symptoms will steadily lose a greater amount of health the more that (s)he is exposed to pollution. For example, an agent with a health of 50 will lose health more rapidly than an agent with a health of 150 when they are equally exposed to over 100µg/m^3 of PM_{10}. Accordingly, the author set the rate of change of an individual's health status caused by PM_{10} exposure to varying linearly with health:

$$If PM_{10} \geq 100, \qquad \frac{dH}{dT} = -\alpha(H_{\max} - H(t)) + H_{\text{recov}} \tag{4.1}$$

, where H_{max} denotes an agent's health status at the beginning and H(t) is strictly less than H_{max}. Thus, in the absence of any recovery and with constant α and PM_{10} always above the threshold, agents' health values would decrease exponentially away from their initial value H(0). H(t) is the current health value. The factor α sets the rate of change per unit of time when the health impact applies. This factor is chosen from a random uniform distribution between zero and a maximum on each tick to allow for the fact that, even within a patch, since these are 30m across, the individual exposure levels will be very different. To some extent, this mimics the fact that people may move in and out of buildings, for example. For vulnerable populations, the first term in equation 4.1 is applied again (so, for example, an agent in the youngest age group has doubled the probability of experiencing an effect). H_{recov} is a health recovery rate that varies by the real estate price of the agent's home location, as in Figure 4.5, up to maximum adaptive capacity (AC, i.e. recovery can only increase health up to some scenario-dependent value). For road patches, a 1.5-fold of background PM_{10} was added to the road patches to elevate the particulate levels near road patches. As with the health loss parameter, parameters for road pollution were also tested in the following section.

Overall, this gives us a threshold model in which impacts begin to accumulate whenever the air quality is marked as unhealthy. However, random variations are applied both to the above agent selection and to the health impact to allow for the fact that the exposure will not be uniform across agents, even when they are more or less spatially co-located (e.g. they may be inside or outside a building or inside a vehicle rather than walking at the roadside).

4.4.7 Scenario Description

This study set three scenarios for pollution trends and three scenarios for adaptive capacity control (see Table 4.2). These were mainly designed as 'what if' scenarios to explore possible consequences in the future. Pollution scenarios were tested as a means of showing how the constant upper or lower trend of PM_{10} can trigger the population's acute illness. Resilience scenarios can be used as an indicator to reflect individuals' health characteristics and the timing of the community response. However, a full

sensitivity analysis is beyond the scope of this paper; in this first glance, this study simply wished to explore the kinds of effects that the above model set-up produces.

Table 4.2 Summary of the policy scenarios

Type	Description			Count
PM Control	BAU	INC	DEC	3
Resilience	100	150	200	3

Pollution Scenarios

The pollution scenarios consist of business as usual (BAU), increase (INC), and decrease (DEC). BAU refers to, 'What if the seasonal pollution levels continue for another period?', which assumes that the six-year time series of hourly PM_{10} in 2010–2015 will replicate itself for another six years. INC projects an upward trend of the seasonal PM_{10} averages by 3% every season, whereas DEC anticipates an equivalent decrease (see Figure 4.9).

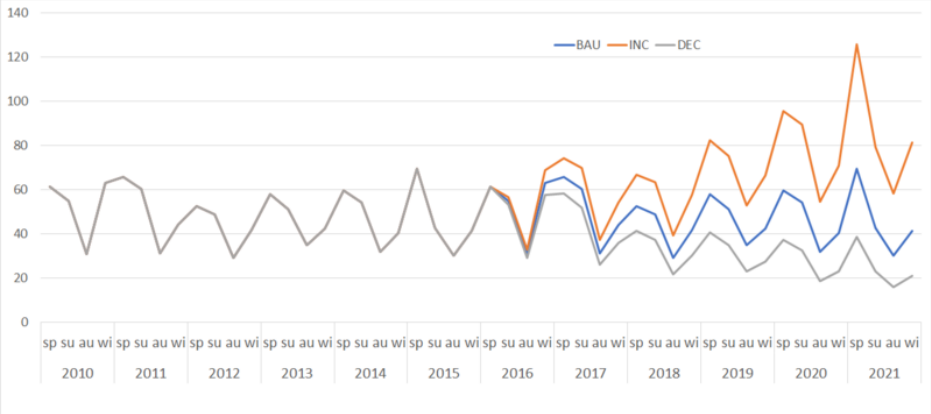


Fig. 4.9 The seasonal PM_{10} mean of Gangnam for the period 2010–2015. Resilience scenarios forecast the next five year trend by replicating this period, referred to as business as usual (BAU), projecting a 3% increase for each season (INC) and a 3% decrease for each season (DEC)

Resilience Scenarios

These scenarios change the 'adaptive capacity' (i.e. the maximum health to which recovery is possible for an agent, hereinafter AC) of 100, 150, and 200. This simulates

mitigating measures, such as improved treatment for the effects of exposure. As a brief illustration, if the resilience is set for AC100, an agent whose health drops below 100 will try to recover up to a maximum of 100.

4.4.8 Model Interface

The model environment was derived from a GIS data set of each district (see Figure 4.10). For simplicity, building and traffic information was excluded for the present. The spatial extent of Gangnam, Mapo, Gwanak, and Jongno, below, are around 40km^2, 24km^2, 30km^2, 24km^2, respectively, with a 30m $\times$ 30m spatial resolution. This study used two time-steps per day (i.e. home hours, working hours), and simulated 12 years, of which the earlier six years of PM_{10} were from the observational data set. For the latter six years, the first six years for which the data was available was reused but was modified to create scenarios for future projections.

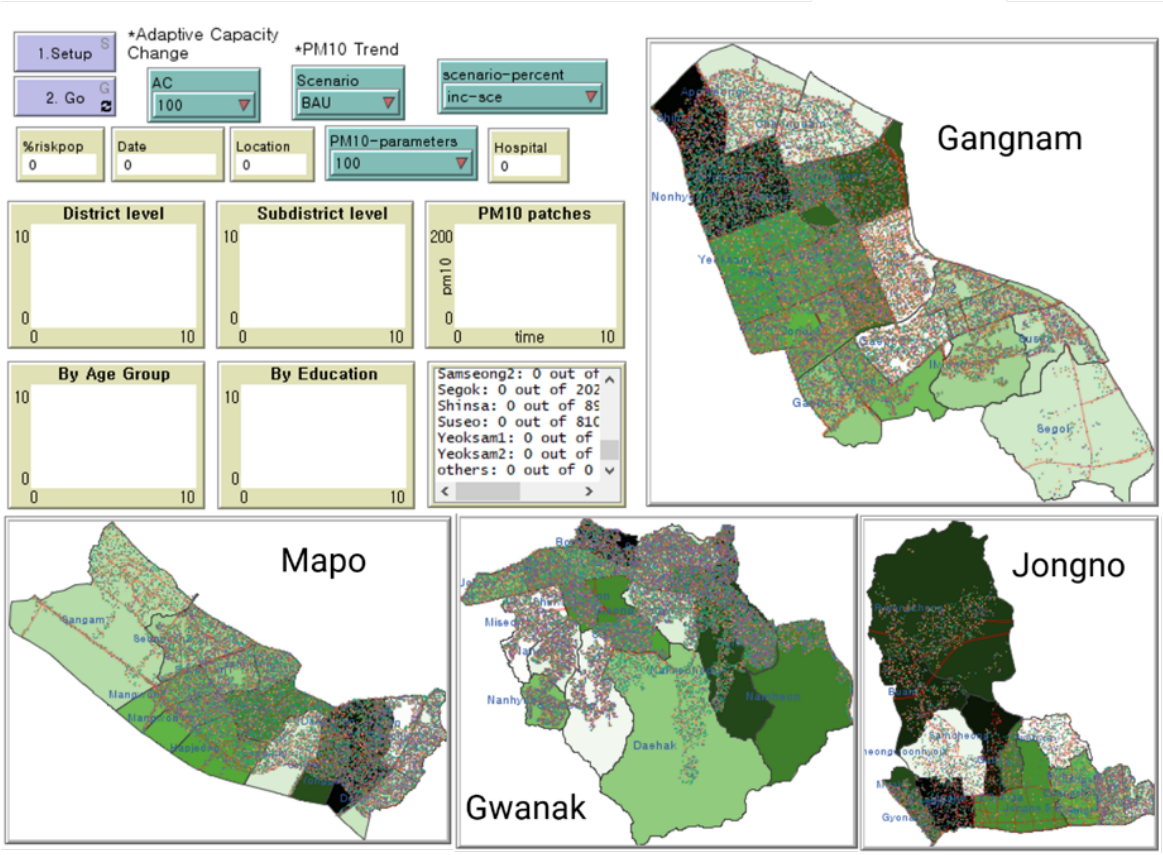


Fig. 4.10 Implementation in Netlogo 6.0.4 for Gangnam, Mapo, Gwanak, and Jongno

4.5 Sensitivity and Calibration

Sensitivity analysis (SA) assesses the behaviour of input parameters that determine the outcome and helps optimise the parameters to make the model resemble the system (Park and Lek, 2016). Here, this study assessed the main factors that differentiate the size of the total at-risk population, then calibrated the population with hospitalised patients. One-factor-at-a-time (OFAT) analysis was applied for each parameter.

4.5.1 Measuring Sensitivity to the Risk Population

Two parameters were included for the sensitivity analysis: health loss (general) and road proximity. Gwanak and Gangnam were chosen because of their similarity in area size but also their distinctiveness in the recovery rate based on land prices. For each implementation, the section plots time on the x-axis and the proportion of the population at risk on the y-axis, where "at-risk"[2] refers to agents whose health status is below 100. The outcome was averaged from 20 iterations to reduce the stochastic variability. Note that the stochastic variability was negligible.

Table 4.3 is a summary of the parameters selected and tested. Health-loss (general) controlled the leverage of health loss across the whole population, while Health-loss (age groups) were used to adjust the values by each age group, but are not shown here because each age group used different parameters in order to fit the hospital admission data. Road proximity was tested from 1, assuming no additional effect from the roads, to 1.5 and 2 times greater PM_{10} concentration than that of the background.

Table 4.3 Demographic parameter values used for calibration and tested for model sensitivity

Parameters	Selected	Tested
Health loss (general)	0.0043	0.001 - 0.01 (by .001)
Health loss (age groups)	Differs by age groups	Numerous combinations
Road proximity	1.5	1 1.5 2

[2]This study uses at-risk as a synonym of health risk, health outcome, population at risk

Figure 4.11 illustrates illustrates the sensitivity of health loss against road proximity. Overall, health loss was sensitive to small unit changes, where a 0.001 rise of α can result in a 10-20% difference to health risk output. Adjusting the parameters from 0.003 to 0.2, the health risk of Gangnam and Gwanak resulted in 0-32% and 0-37%, and both districts showed a 16-18% gap of at-risk rates between 0.005 and 0.006. Notably, the timing of health risk surge varied by districts but also the road proximity parameters. The tipping point between districts was mainly due to the 15-64 age group whose health went below 100. Road proximity affects the outcomes to change linearly as the parameters increased.

Having tested the combinations between both parameters and value adjustments, this study selected health loss at 0.0043 and road proximity at 1.5 to be implemented in scenarios, with the evidence that PM_{10} on roads are approximately 50% higher on average than the background concentrations (from the data exploration in Section 3.3.2 that compared background and roadside station). The final parameter value resulted from the calibration with the hospital admission data from HIRA (Health Insurance Review & Assessment Service). The processes will be further explained in the next section. Note that the selected parameter values are all illustrative and it can depend on the adjustment of other settings, e.g. calibration data.

The onset of health risk, the first model tick at which the total percentage population at risk exceeded 0%, was also experimented with using health loss parameters. In Figure 4.12, the onset appeared after 6000 ticks in Gangnam and 5300 ticks in Gwanak. Despite a small difference of time between districts, both districts saw decreases as the parameter incremented, and the curve seemingly begins to flatten after 0.004.

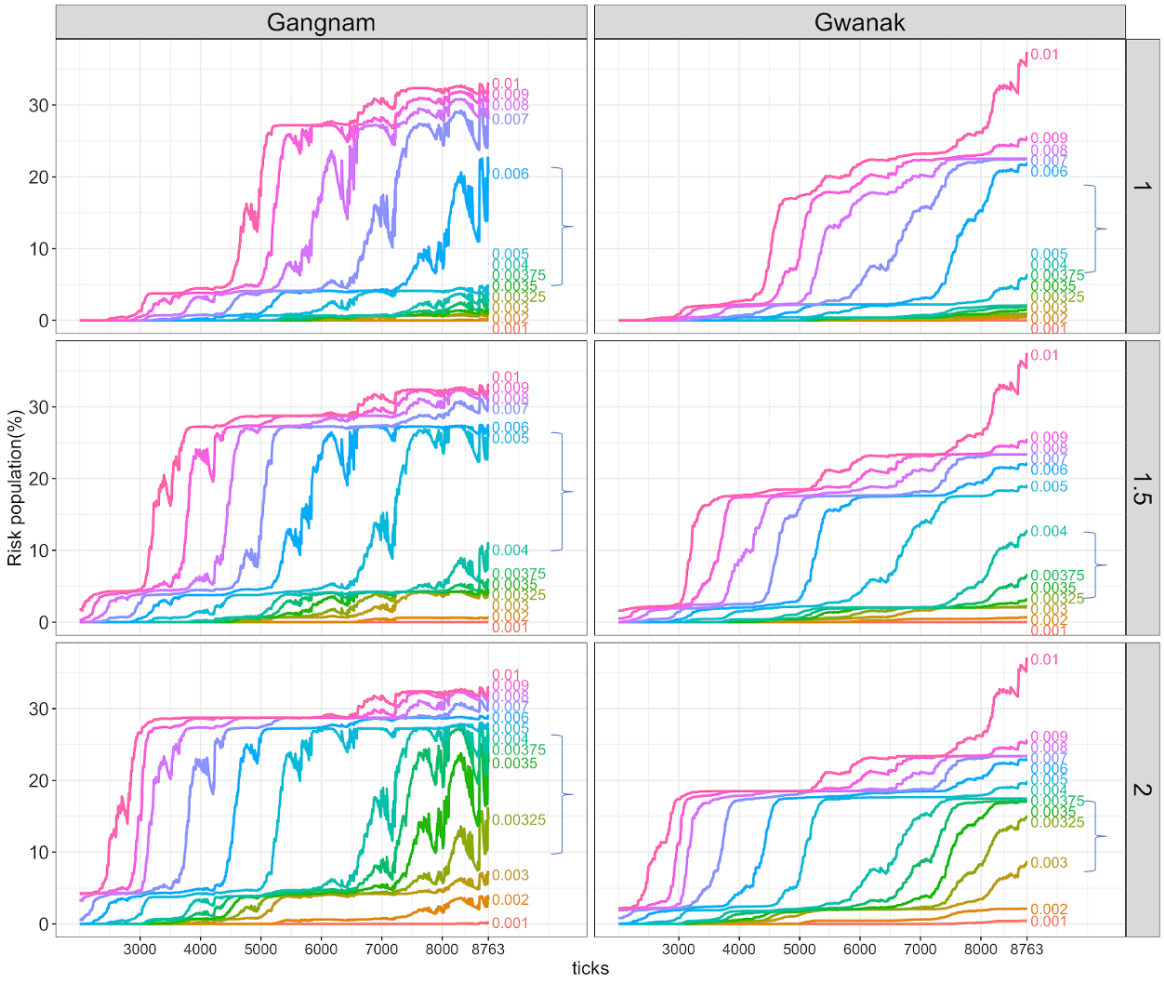


Fig. 4.11 Sensitivity of health loss and road proximity by time and risk population. The 3 by 2 array is indexed by districts and pollution weights to road patches. The curly brackets show the decision threshold. In both districts, the decision threshold lies between .003 - .005

4.5.2 Calibrating Unwell Agents to CDC Patient Data

Data Description

The model counted those whose health is 0 and under as in hospital for treatment. Hospital patient data were made available through the Korean CDC (Center for Disease Control and Prevention) database, referred to as HIRA. Within the database, respiratory patients (section J01-J99) of 2016 were selected, excluding J00 (common cold) and patients treated by medical herbalists (acupuncture, moxibustion and so on).

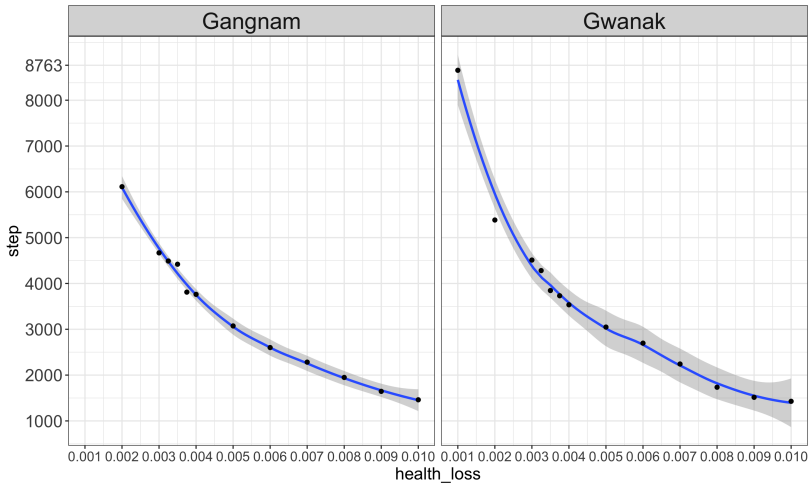


Fig. 4.12 The onset of the first step against health loss parameters in Gangnam and Gwanak. The decreasing trend shows that the curve starts to flatten after 0.004 in Gangnam and 0.003 in Gwanak

The data cost approximately £200 per year and was only given for non-commercial purposes. Since the home address for each patient was anonymised (legal protection of privacy), this study summed patient statistics of all Seoul hospitals by age groups in five-year intervals e.g. 5-9, 10-14. The total number of 230,962 patients from each age group was reduced to a fraction of 5% to match the outcome of the simulation that used 5% of the census. Finally, patients of each district simulated from the BAU-AC100 scenario were added and compared with the final figure (see Table 4.4).

A disclaimer is that the number of patients can vary by each run. The uncertainty of the movement of agents and changes in pollution fields will affect the number of people admitted to hospital.

Table 4.4 5% sample population and patients by age groups from Seoul hospital data 2016

Age	Census 5%	Patients 5% (2yr mean)
05-09	18,663	3,329
10-14	20,249	935
15-19	27,557	1,150
20-24	35,597	1,264
25-29	38,283	1,545
30-34	42,871	2,217
35-39	39,671	1,832
40-44	41,844	1,675
45-49	41,210	1,678
50-54	40,320	2,070
55-59	39,192	2,402
60-64	29,121	2,164
65-69	22,274	2,307
70-74	17,056	2,211
75ov	21,720	2,750

Calibrated results

From 25 district outcomes of the BAU-AC100 scenario, the total number of patients from the model made a reasonable approximation to the observed data (see Table 4.5 and Figure 4.13). Most of the group had a <100 difference between the modelled and observed value. It was possible to reduce the difference between model and observation to less than some percentage of the total (max difference was observed in the 45-49 aged group: 248/1430=17%), and that the variance in the model runs was typically a similar order of magnitude.

The outcomes from the model run revealed that 11 out of 25 districts had patients admitted to the hospital while the others did not. The selected districts include Dongdaemoon, Gangseo, Guro, Gwanak, Gwangjin, Jung, Jungnang, Mapo, Seocho, Yangcheon, and Yeongdeungpo, however, the majority of the patients were found from Gwanak and Mapo.

Table 4.5 The sample number of respiratory patients, modelled results, and their difference

Age	Observed	Modelled	Difference
05-09	3,329	3,370	-41
14-10	935	955	-20
15-19	1,150	1,074	76
20-24	1,264	1,300	-36
25-29	1,545	1,590	-45
30-34	2,217	2,150	67
35-39	1,832	1,628	204
40-44	1,675	1,652	23
45-49	1,678	1,430	248
50-54	2,070	1,827	243
55-59	2,402	2,305	97
60-64	2,164	2,010	154
65-69	2,307	2,083	224
70-74	2,211	2,102	109
75ov	2,750	2,669	81

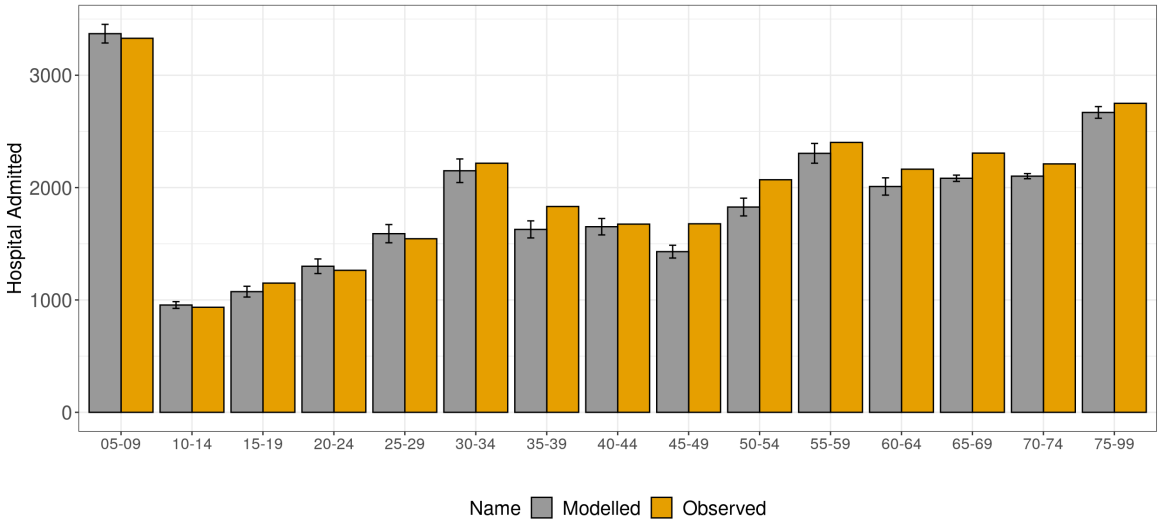


Fig. 4.13 Calibration result of the modelled result against the patient data. The observation data results from a two year average of the Korean CDC 2015-2016

4.6 Simulation Results

4.6.1 A comparison of health vulnerability across 25 Seoul districts

To reiterate, each district has a combination of three categories of pollution scenarios (BAU, INC, and DEC) and three resilience scenarios (AC100, AC150, and AC200). Each scenario plotted time on the x-axis and the proportion of the population at risk on the y-axis.

To avoid the stochastic variability, model runs were averaged from 20 model runs. The chapter discovered that the stochastic variability between model runs was not as large and the average values converged after five runs. The results section presents (i) health vulnerability by demographic factors, (ii) a comparison of health outcomes by districts, (iii) scenario forecasting by districts and (iv) by subdistricts.

Health Vulnerability by Demographic Factors

Overall, the outcomes showed that the health risk (or at-risk) rates were distinctive between the demography subgroups. The elderly were the most sensitive subgroup which ranged between 10-24% in the BAU scenario across districts, then surged up to 60% in the INC; however, it only managed to reduce by a few percentages in the DEC. The 15-64 group also ranged between 2-30% in the BAU across districts but the oscillations (rise of risk vs health resilience) seen in most of the districts, which were different from the outcome of the elderly. In the INC, the at-risk rate of all districts exceeded 20% with a linear increase, meaning that the health risk far outweighed the resilience. Unlike the elderly, the group showed a substantial decrease in the DEC ranging from 0-9%. The under 15s were the group that showed the least difference between scenarios. Despite a 10% range observed between districts in the BAU scenario (4-15%), the at-risk rate in INC and DEC were 11-15% and 2-11% respectively.

Health risks by education were less distinctive compared to the age group. In BAU, the health risk in low and high education groups ranged between 4-25% and 3-31% respectively, INC ranged between 14-38% and 20-36% respectively, and DEC between 3-12% and 1.5-5% respectively. Although a small difference was identified between the education groups, it turned out that, from a temporal perspective, there was no clear

association with education (see Figure 4.2 to examine the distribution by education and age groups).

Figure 4.14 illustrates the distribution of health vulnerability by districts in Seoul (also see the district codes in Table 4.6). Note that the percentage of the age maps sum up to the total and that of education levels sum up to the total respectively, but the sum of the age and education subgroups does not match the total. Using the outcome of BAU×AC100 as a default structure, the highest risk areas included Yeongdungpo, Mapo, and Guro, which had over 20% of the at-risk rate. The total percentage was greatly affected by the 15-64 age group ('economically-active') because the group accounted for the majority of the city's population. According to the outcomes, the risk rates of high education seemed to have a high correlation with the economically active group. This can be seen through the colour codes between the maps. Although it is not presented in the figure, the study sampled the agents in a few districts and found that those with high health risks in the economically active group were most likely to have higher education status.

In the regional aspect, there was a huge contrast between the central west with a higher risk rate and the northern districts with a remarkably low risk rate. The low-risk areas were consistent across the maps. However, the outcome of the elderly was comparable between districts.

Table 4.6 District name and code used in Figure 4.14 maps

District	Code	District	Code
Dobong	DBG	Jung	JUN
Dongdaemoon	DDM	Jungnang	JNG
Dongjak	DON	Mapo	MAP
Eunpyeong	EUN	Nowon	NOW
Gangbuk	GBK	Seocho	SCO
Gangdong	GDG	Seodaemoon	SDM
Gangnam	GNM	Seongbuk	SBK
Gangseo	GSO	Seongdong	SDG
Geumcheon	GEU	Songpa	SPA
Guro	GRO	Yangcheon	YCN
Gwanak	GWK	Yeongdeungpo	YDP
Gwangjin	GWJ	Yongsan	YSN
Jongno	JNO		

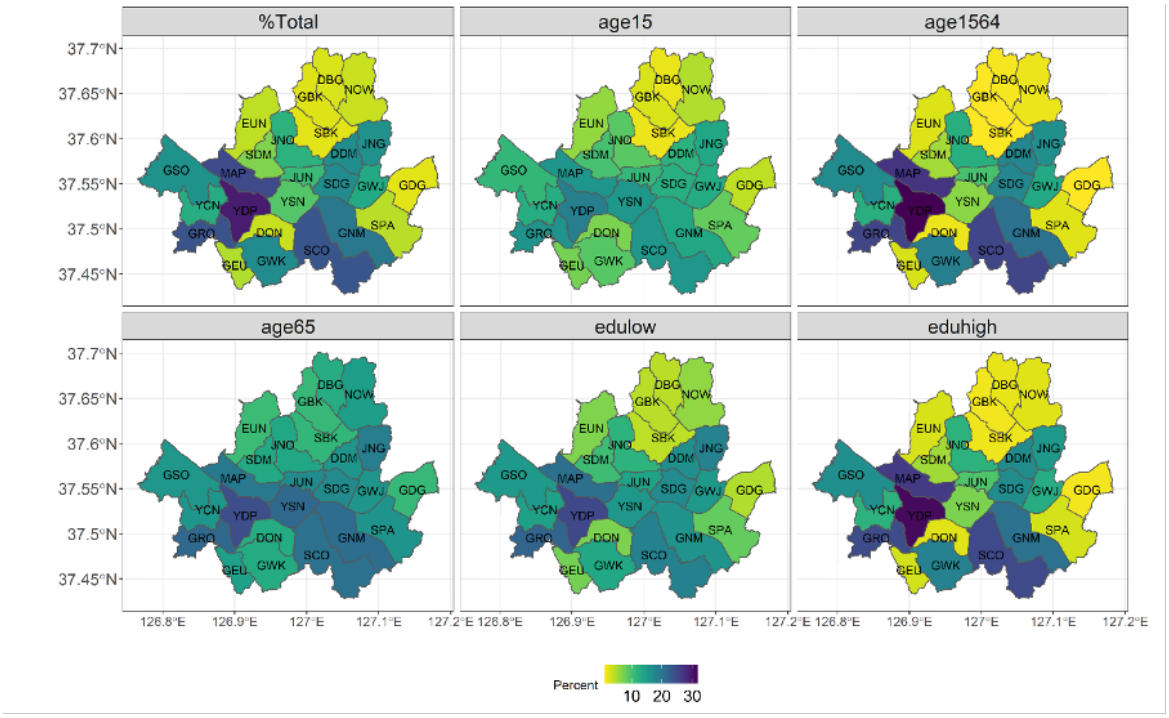


Fig. 4.14 Health risk in Seoul Districts as a result of BAU×AC100. The districts used abbreviated codes to avoid any confusion.

BAU Scenario

This section presented the outcomes from the combination of BAU and three health resilience scenarios, AC100, AC150, and AC200. To reiterate, the BAU scenario assumed that the six years of PM_{10} records were to be replicated for another six years, and the health resilience (i.e. adaptive capacity) assumed different levels of which an individual's nominal health can recover.

In the BAU×AC100 (see Figure 4.15-Top), the at-risk population in most districts started to increase between 4500 and 6000 ticks, and from 7200 ticks to the end of the simulation. This might be expected as maximum health was equally given to agents at the beginning of the simulation, and similar health loss happened across agents in the same group as PM_{10} generated from each district monitor gave similar exposure levels. When health resilience increased from AC100 to AC200, the at-risk population over 10% declined below 5% in most of the districts.

Additionally, the time-series plots showed an oscillated trend. This may have resulted from the balance between the exponentiality of health loss and the linearity

of resilience from the resilience scenario and the land price effect. The districts that had oscillated trends of risk population were mostly in the AC100 scenario, between age 15 and 64, and both education groups. These groups seemed to have been associated with the balance between recovery rate and degrading of health with exposure (see Figure 4.16). The districts that did not have a repetition of oscillation had been continuously exposed to extreme PM_{10} episodes, which made health recovery ineffective. For the elderly group, the oscillations only appeared in a few districts, but these districts experienced a linear upward trend due to the weak capacity of health that the group can recover.

With regard to age and education, people aged 15–64 ("economically active") as well as highly educated faced the highest risk in AC100 but decreased significantly to be the lowest risk of the population in AC200, i.e. high resilience scenario (see Figure 4.16). This might be expected from the model setup: one could imagine a case in which the active population were more exposed as a result of travel on very polluted highways, or being outdoors in polluted air more, but tends not to be the case when the resilience remains high. Barely anyone in this group showed illness until 4500 ticks when the first phase showed around 3% risk population in most districts, then over 10 districts soared by 10% after 7000 ticks being identified as the second phase. At the final tick of the AC100 scenario, 10 districts exceeded 15% which include Yeongdeungpo, Mapo, Guro, Seocho, Gangnam, Gwanak, Dongdaemoon, Seongdong, Gangnamseo, and Jungnang. This rate dropped significantly in the AC200 scenario by <5%.

By contrast, those aged over 65 experienced risks to health earlier than the young and less educated, and did not show effective signs of improvement when higher resilience of health was applied. In AC100, the top districts which had the highest health risk included Yeongdeungpo (24.3%), Guro (21.5%), Yongsan (21.3%), Gangnam (20.9%), and Seocho (20.1%), which were all over 20%, while the lowest district was Eunpyeong (10.8%). However, the health risk was maintained regardless of resilience scenarios. This can be seen in the AC200 results of Yeongdeungpo (21.6%), Guro (19.1%), and Yongsan (17.1%), where they only had 3% difference when higher health resilience was applied in the model. The under 15s had the lowest risk percentage of all age groups. Yeongdeungpo had the highest risk of all with 18.6% for AC100, followed by Guro (16.0%), Mapo (15.6%), Yeongsan (15.5%), and Seocho (15.4%). Health risk decreased by 6-7% in AC200.

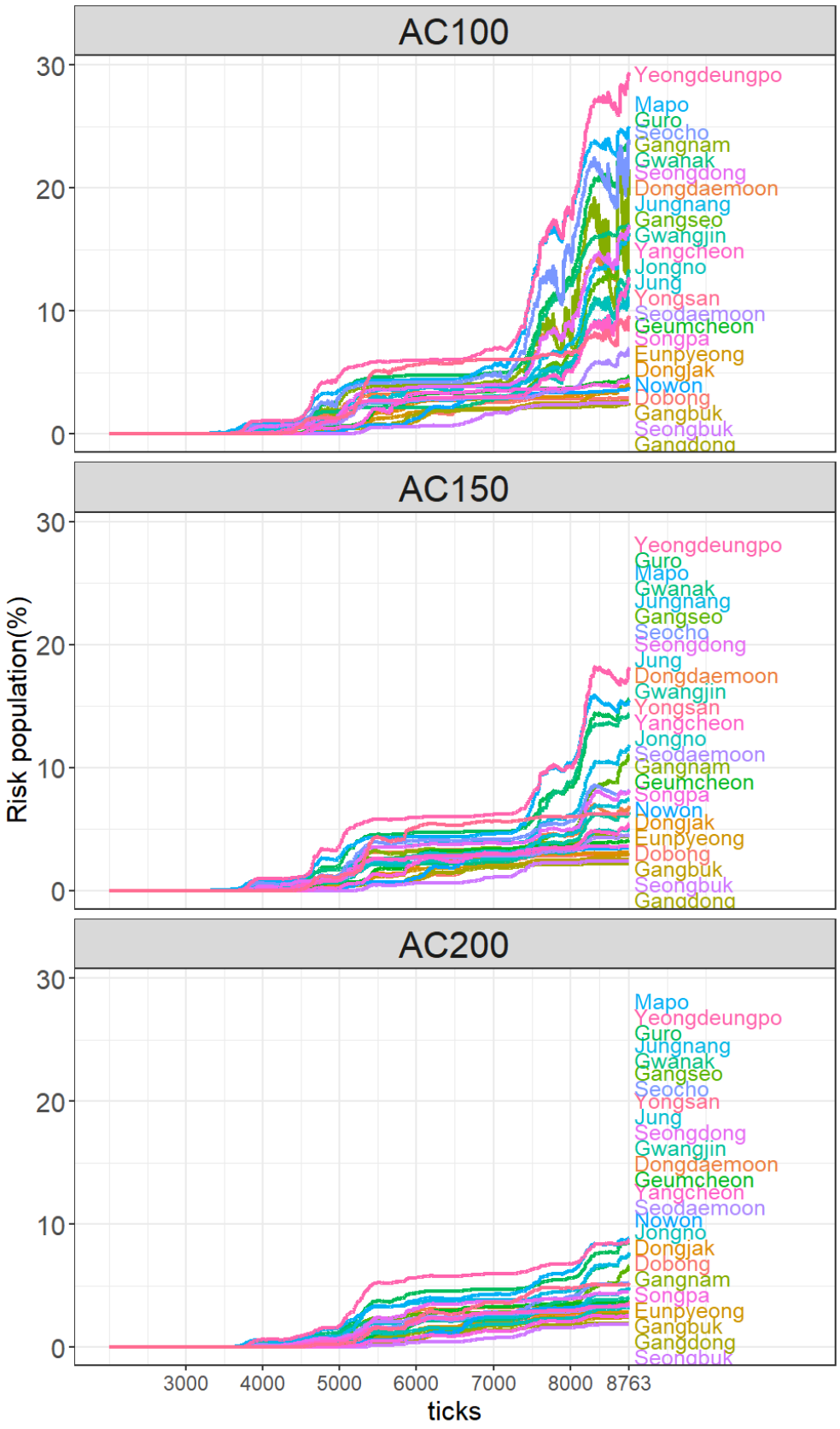


Fig. 4.15 The risk population change amongst 25 districts in BAU scenarios. The columns are grouped by resilience scenarios (i.e. AC100, AC150, and AC200) and the rows are grouped by the total risk change

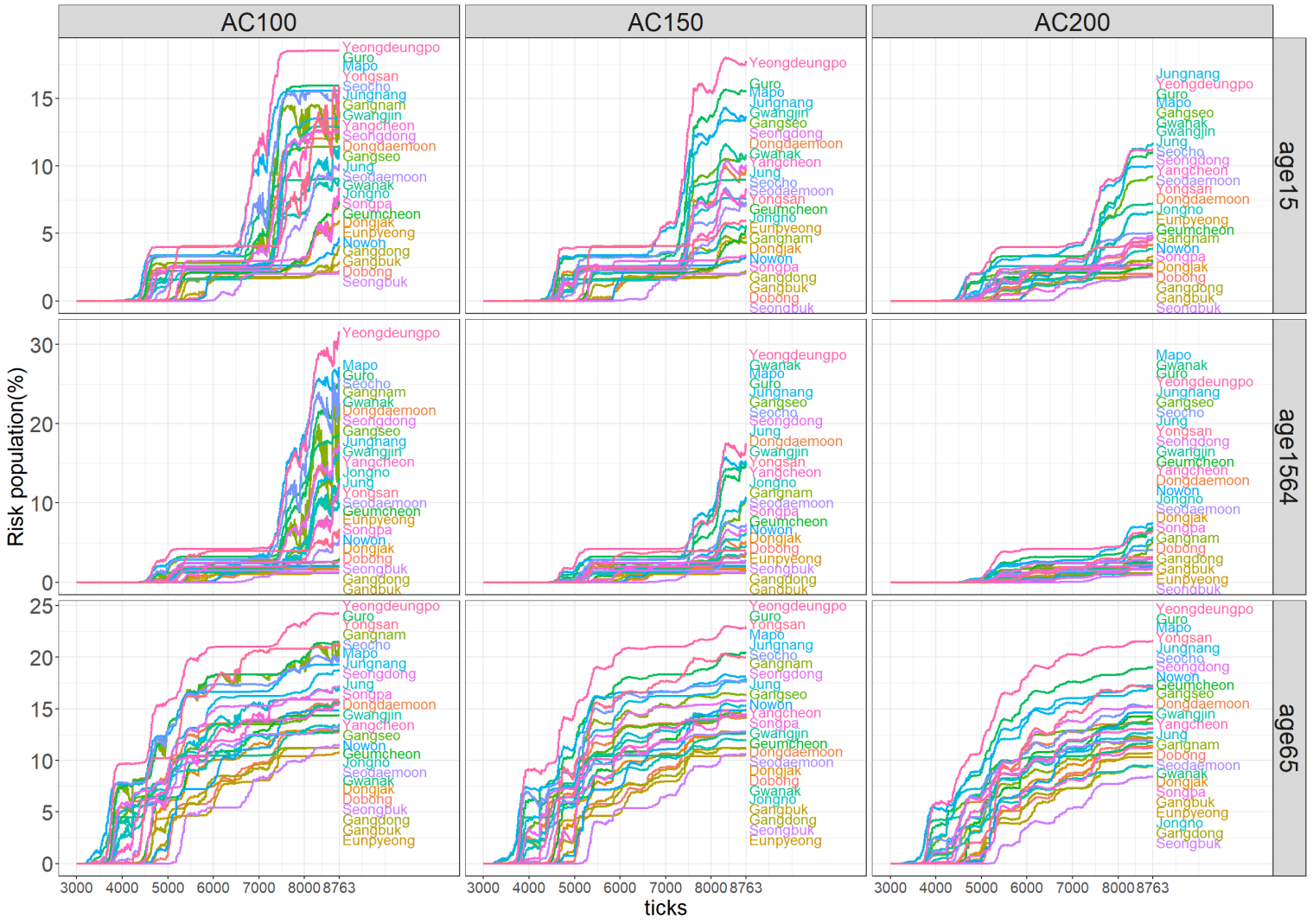


Fig. 4.16 The risk population change amongst 25 districts in BAU scenarios by age group. The columns are grouped by resilience scenarios (i.e. AC100, AC150, and AC200) and the rows are grouped by age groups.

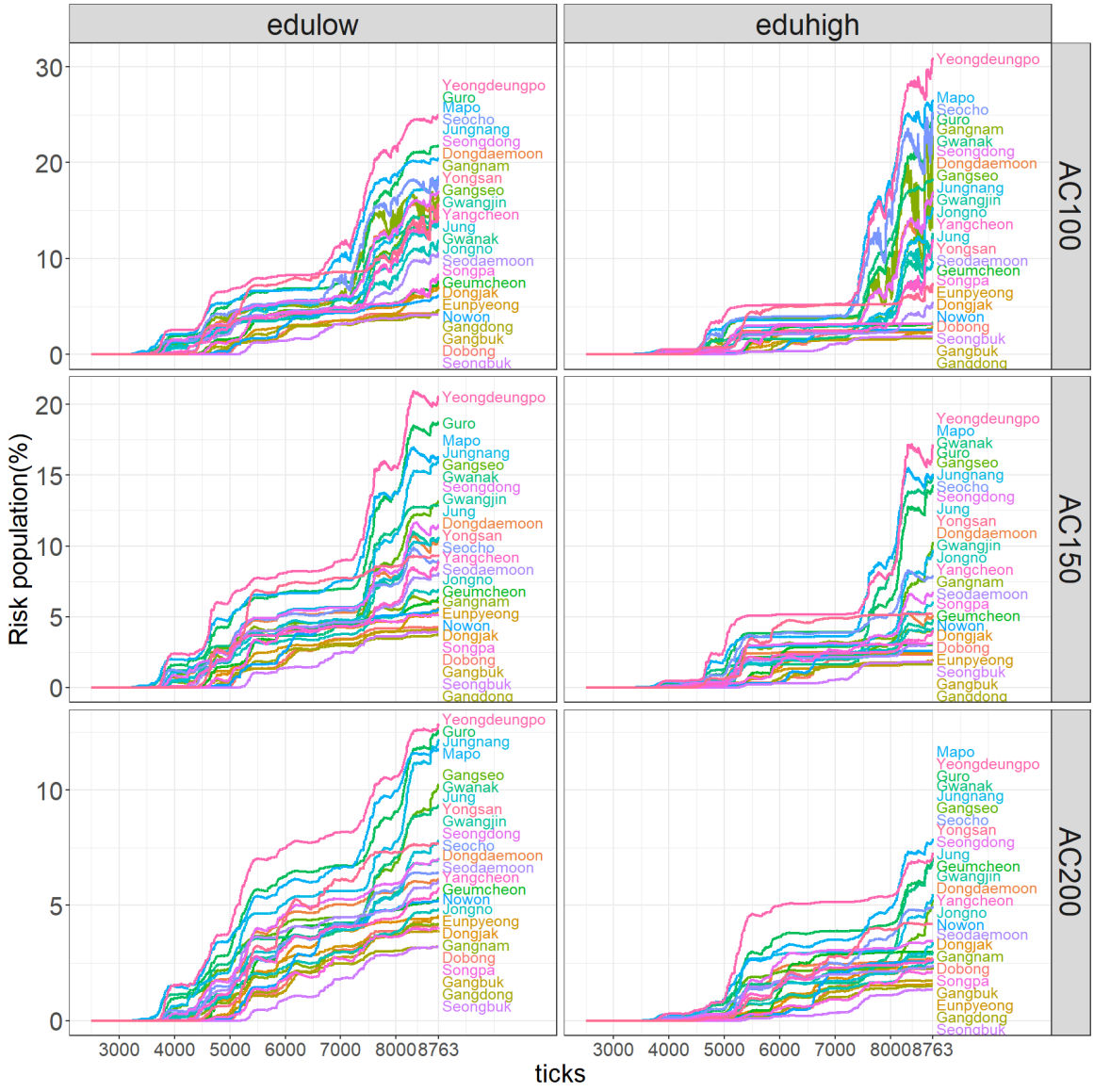


Fig. 4.17 A comparison of the risk population change between 25 districts in BAU scenarios. The columns are grouped by resilience scenarios (i.e. AC100, AC150, and AC200) and the rows are grouped by educational status.

Table 4.7 Population at risk by districts and resilience scenarios and the difference on the final tick in the BAU Scenario

District	AC100	AC150	AC200
Dobong	2.9	2.9	2.8
Dongdaemoon	16.8	6.8	3.8
Dongjak	3.7	3.1	2.9
Eunpyeong	4.1	3	2.5
Gangbuk	2.7	2.6	2.5
Gangdong	2.5	2.2	2
Gangnam	19.4	4.4	2.7
Gangseo	16.2	11	6.6
Geumcheon	4.7	4.1	3.7
Guro	23.8	15.6	8.7
Gwanak	17	14.4	7.6
Gwangjin	13.3	6.4	4.1
Jongno	12.1	5.1	3.3
Jung	11	7.5	4.8
Jungnang	16.3	11.8	7.7
Mapo	24.8	15.3	8.9
Nowon	3.7	3.5	3.4
Seocho	23.5	8.1	5.3
Seodaemoon	7	4.6	3.4
Seongbuk	2.5	2.4	1.9
Seongdong	17	8.1	4.5
Songpa	4.4	3.6	2.6
Yangcheon	12.7	5.5	3.6
Yeongdeungpo	29.3	18.1	8.8
Yongsan	9.3	6.3	5.1

INC Scenario

From Figure 4.18-Figure 4.20, the risk population in the INC scenario, which assumes a 3% increase in PM_{10} per season starting from the seventh year, showed a small rise around 5% by 3500 ticks but showed a steep incline to 30% after 7000 ticks. Yeongdeungpo had the highest risk population with 37.8% and the average risk population was 27.8%. The highlighted districts in the age and education plots are the ones which exceeded 30% of risk population.

It was notable that the 'stepped'-type increase, also known as a reoccurrence of the "increase-stable-increase-stable", only happened in the INC scenario. The initial rise of health risk was found from the elderly and the young, and then the next rise occurred from the economically active. The study did not observe any sign of health resilience both individually and regionally because the 3% seasonal rise of PM_{10} outweighed the resilience. This implies that as more days are projected to be polluted, the vulnerable groups, particularly the elderly, will expect to be at risk regardless of health care support, followed by other groups of the population which reside nearer to roads.

Nevertheless, the temporal trend of the at-risk population varied by demographic group. While over 65s had experienced a constant increase until 7500 ticks, then a final surge at around 75%, the other two groups showed a similar pattern that showed an initial onset at 4300 ticks, plateaued for a while, and surged in the final quarter of the simulation. Much of this happened because the elderly did not have enough health resilience against extreme PM_{10} that had shown in the BAU scenario outcome, whereas the other two groups had different time (ticks) of onset but had enough resilience of health that maintained the at-risk level to some extent. However, the reassuring fact was that the majority of the youth and the economically active managed to prevent health deterioration from the consistent risk of PM_{10}.

There was no clear relationship between health risks and education (see Figure 4.19). However, this study presumed that the variety of education levels seen in the economically active group has varied the outcome (see Figure 4.2 for the distribution), while the under 15s and over 65s which comprise less than a quarter of the population were more likely to have low education because the youth are still in their learning process and have not met the criteria and the elderly (born before the 1950s) may have missed the chance of receiving an education.

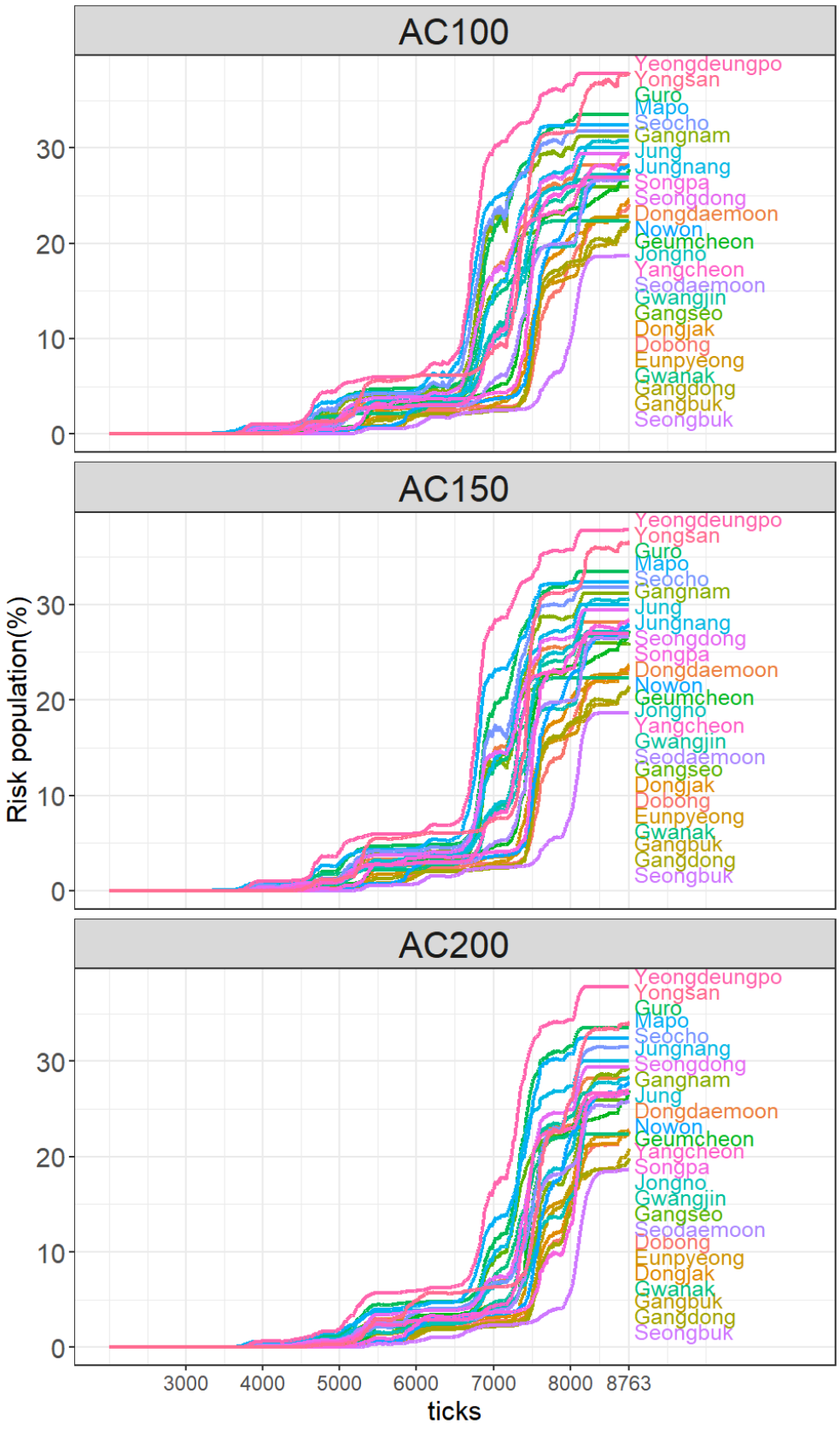


Fig. 4.18 The risk population change amongst 25 districts in INC scenarios. The columns are grouped by resilience scenarios (i.e. AC100, AC150, and AC200) and the rows are grouped by the total risk change

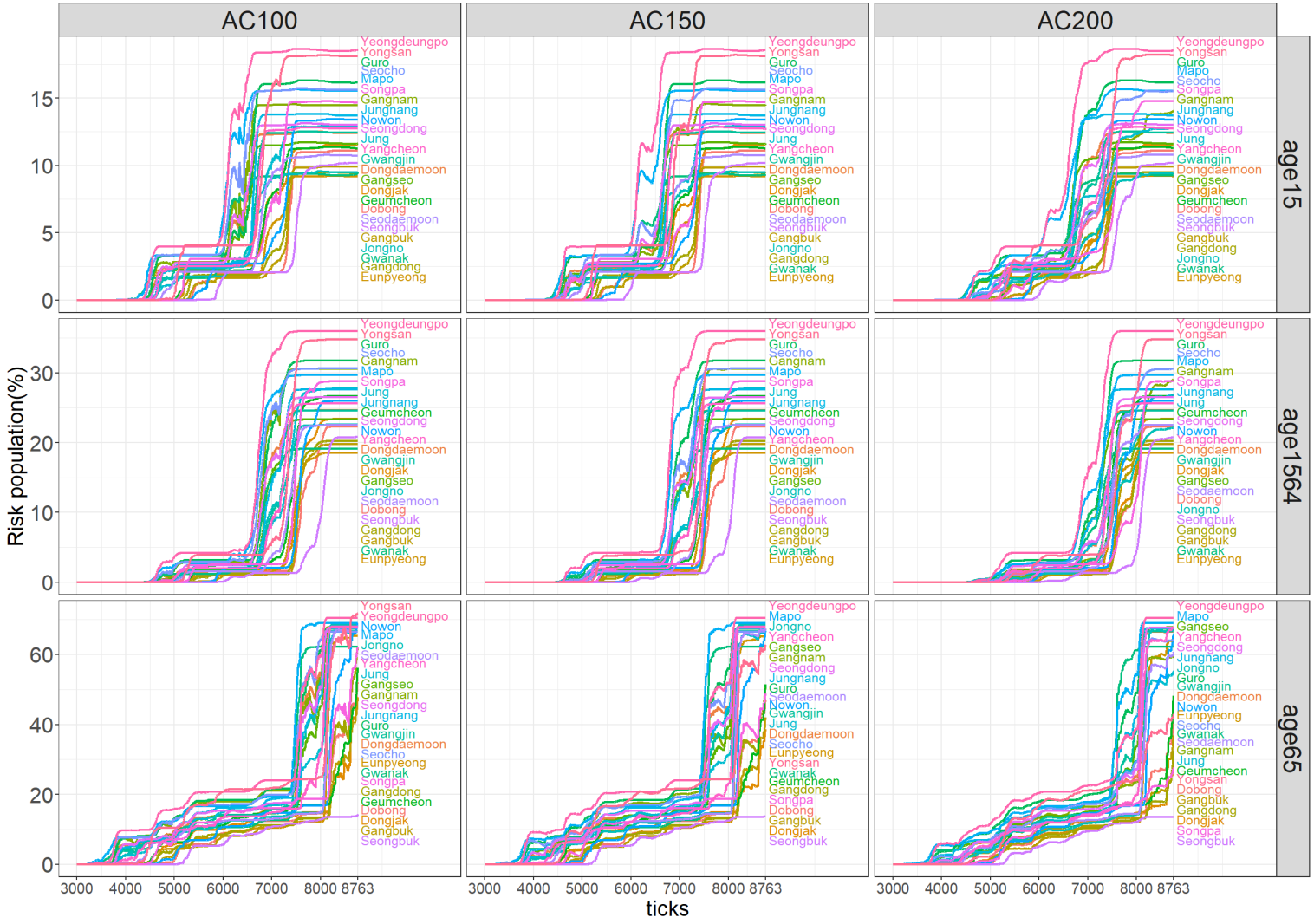


Fig. 4.19 The risk population change amongst 25 districts in INC scenarios by age group. The columns are grouped by resilience scenarios (i.e. AC100, AC150, and AC200) and the rows are grouped by age groups.

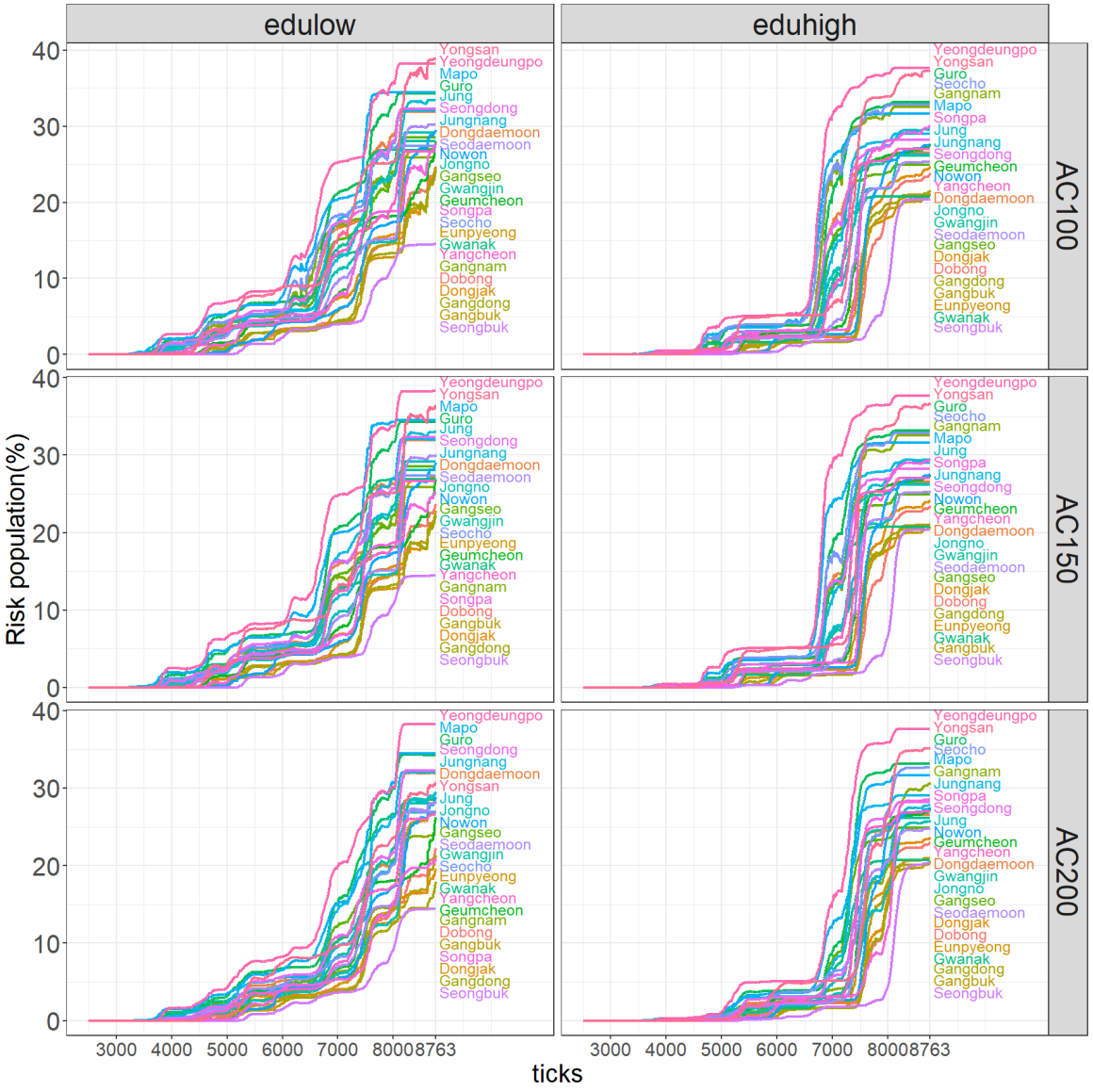


Fig. 4.20 A comparison of the risk population change between 25 districts in INC scenarios. The columns are grouped by resilience scenarios (i.e. AC100, AC150, and AC200) and the rows are grouped by educational status.

Table 4.8 Population at risk by districts and resilience scenarios and the difference on the final tick in the INC Scenario

District	AC100	AC150	AC200
Dobong	24	23.5	22.8
Dongdaemoon	28.2	28.2	28.2
Dongjak	24.6	23.7	22.6
Eunpyeong	22.8	22.8	22.7
Gangbuk	22.1	21.4	20.8
Gangdong	22.3	21.4	19.8
Gangnam	31.2	31.2	29.3
Gangseo	25.9	25.9	25.9
Geumcheon	27.8	27.3	27
Guro	33.5	33.5	33.5
Gwanak	22.4	22.3	22.3
Gwangjin	26.7	26.7	26.7
Jongno	27.2	27.2	26.7
Jung	30.8	30.6	28.4
Jungnang	30	30	30
Mapo	32.4	32.4	32.4
Nowon	28.2	28	27.9
Seocho	31.8	31.8	31.5
Seodaemoon	26.7	26.6	25.8
Seongbuk	18.8	18.7	18.7
Seongdong	29.4	29.4	29.4
Songpa	29.5	28.4	26.8
Yangcheon	26.9	26.9	26.9
Yeongdeungpo	37.8	37.8	37.8
Yongsan	37.7	36.6	34

DEC scenario

In the DEC scenario (see Figure 4.21 - Figure 4.23), which assumes a 3% decrease in the PM_{10} per season starting from the seventh year, the risk population in the oscillated group fluctuated in all of the resilience scenarios.

Overall, despite the oscillations appearing in several districts, health risks in most districts stayed under 5% in AC100, and the trend maintained across higher resilience scenarios with less oscillations. However, for a few districts that include Dongjak, Gangnam, Songpa, and Seongbuk, they showed a remarkable decrease that brought the health risk down to <1% in AC200, implying almost all residents are not at risk.

Unlike the overall outcome, health risk by demographic groups appeared surprisingly different. While the health risk of the young and the economically active that had continuous oscillations after the action of onset in AC100 was flattened in higher resilience scenarios (AC200), the elderly only experienced a marginal decrease. This reveals that although PM_{10} is projected to decline every season and the effect of health resilience is provided, the elderly can still get adverse health effects attributed to PM_{10}. As with the previous section, there were hardly any relationships shown in the education results because of the variety of education levels seen in the economically active group (see Figure 4.23).

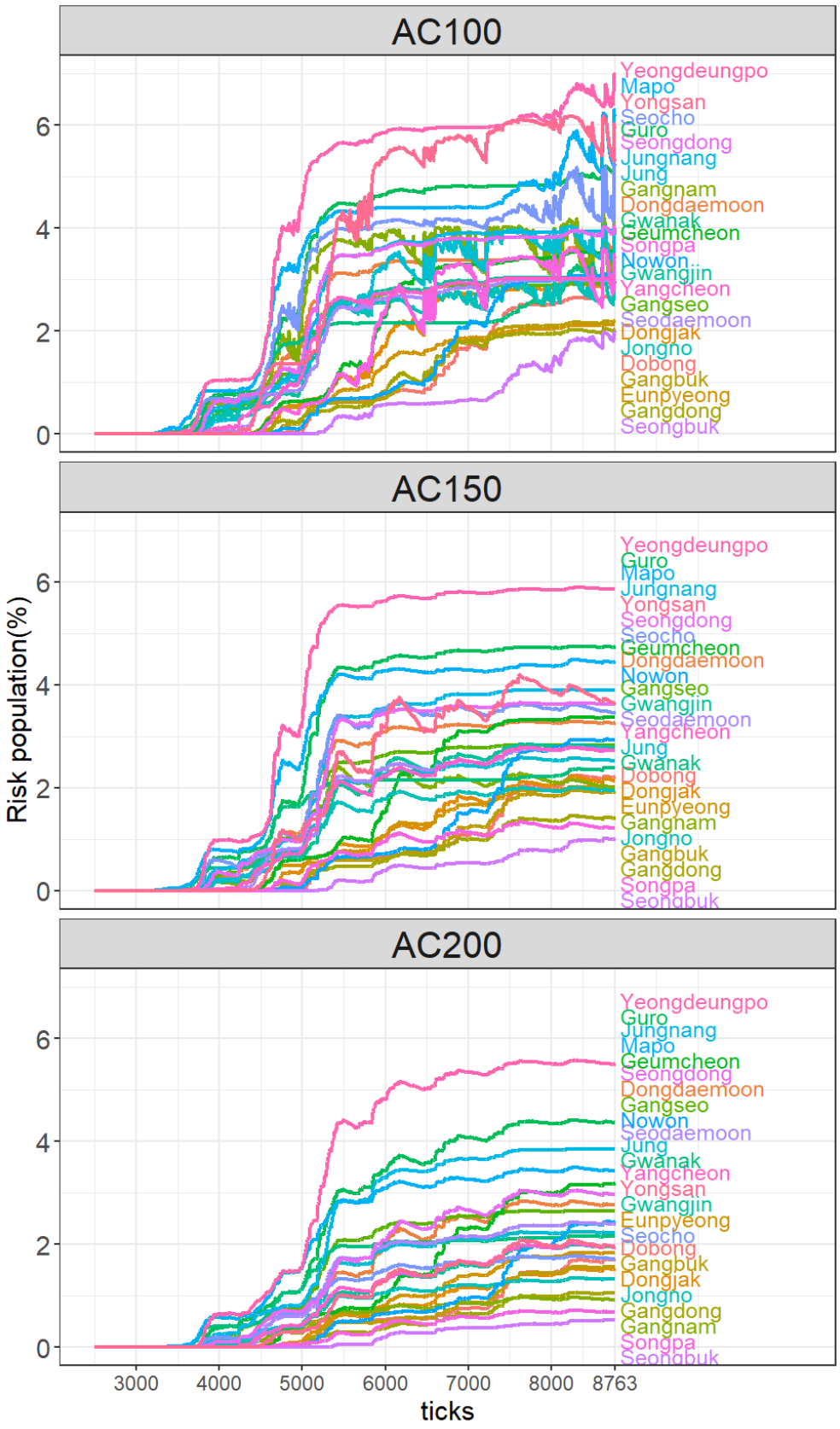


Fig. 4.21 The risk population change amongst 25 districts in DEC scenarios. The columns are grouped by resilience scenarios (i.e. AC100, AC150, and AC200) and the rows are grouped by the total risk change

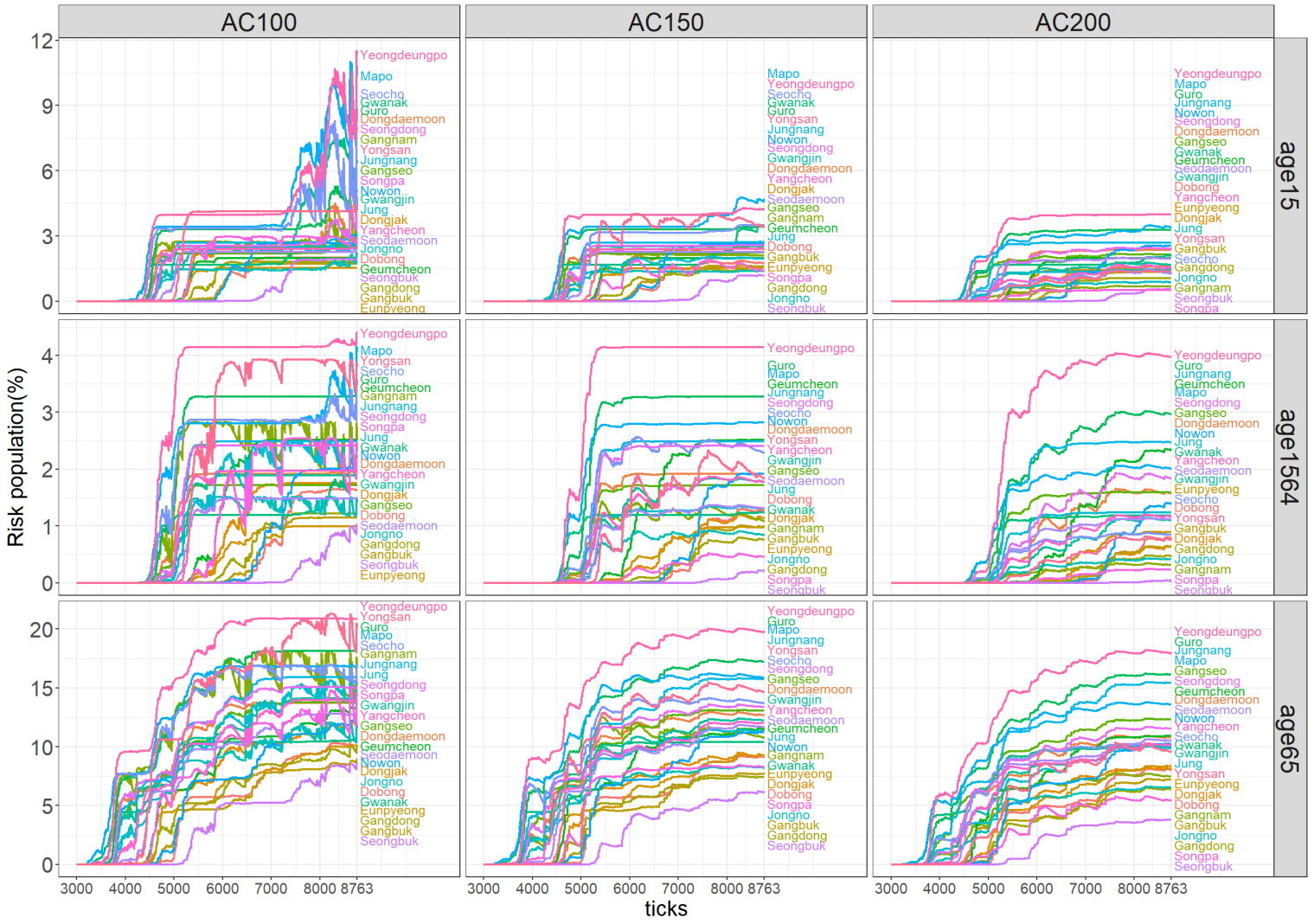


Fig. 4.22 The risk population change amongst 25 districts in DEC scenarios by age group. The columns are grouped by resilience scenarios (i.e. AC100, AC150, and AC200) and the rows are grouped by age groups.

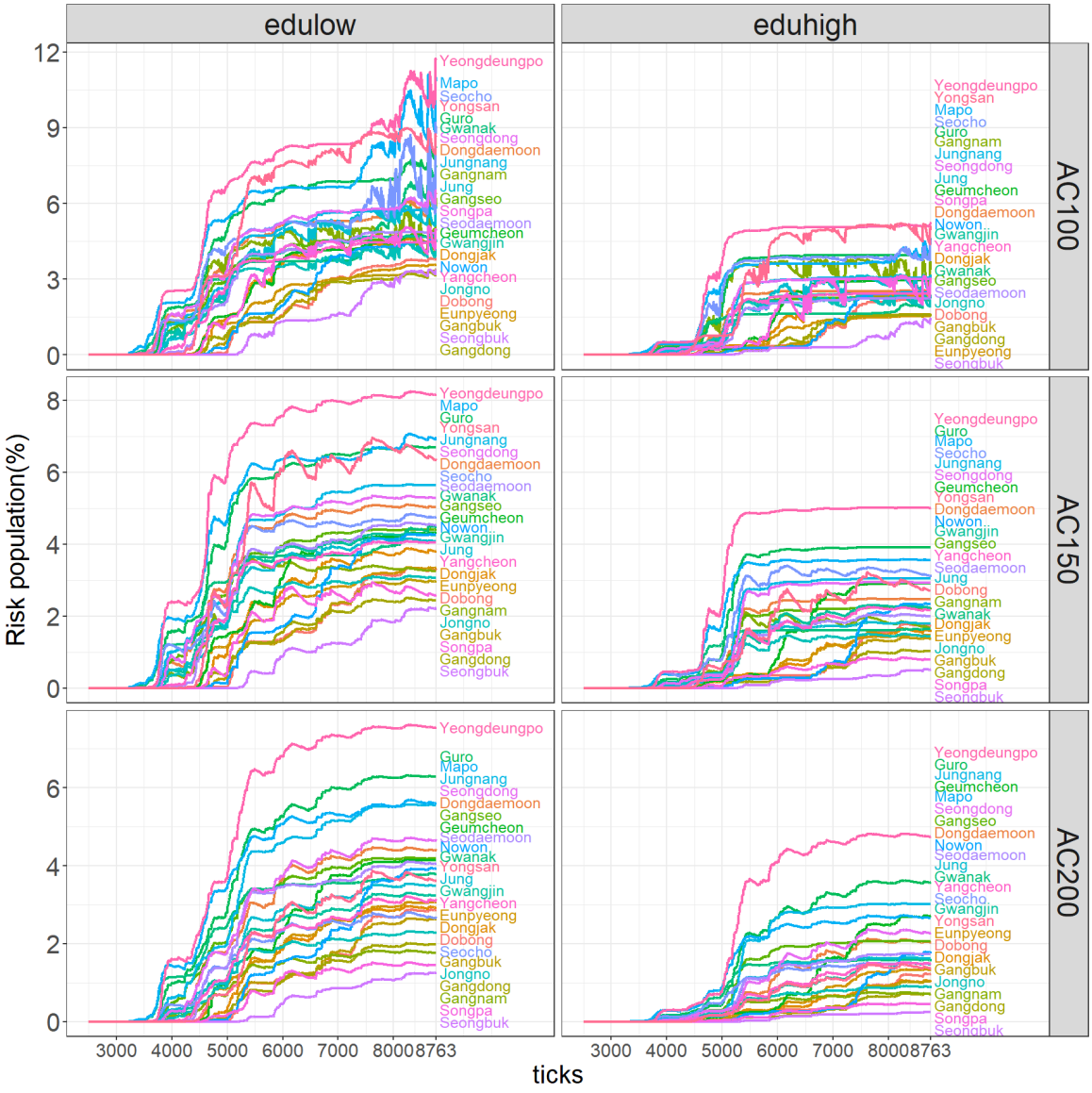


Fig. 4.23 A comparison of the risk population change between 25 districts in DEC scenarios. The columns are grouped by resilience scenarios (i.e. AC100, AC150, and AC200) and the rows are grouped by educational status.

Table 4.9 Population at risk by districts and resilience scenarios and the difference on the final tick in the DEC Scenario

District	AC100	AC150	AC200
Dobong	2.7	2.2	1.7
Dongdaemoon	3.7	3.3	2.8
Dongjak	2.9	2.2	1.5
Eunpyeong	2.1	2	1.8
Gangbuk	2.2	1.9	1.6
Gangdong	2	1.4	1.1
Gangnam	3.9	2	0.9
Gangseo	3	2.8	2.6
Geumcheon	3.6	3.4	3.2
Guro	5.2	4.7	4.4
Gwanak	3.6	2.4	2.2
Gwangjin	3.1	2.8	1.9
Jongno	2.9	1.9	1.3
Jung	3.9	2.5	2.2
Jungnang	4	3.9	3.9
Mapo	6.2	4.4	3.4
Nowon	3.1	2.9	2.4
Seocho	5.3	3.5	1.7
Seodaemoon	3	2.7	2.4
Seongbuk	2	1	0.5
Seongdong	4	3.6	3
Songpa	3.4	1.2	0.7
Yangcheon	3	2.7	2
Yeongdeungpo	7	5.9	5.5
Yongsan	5.9	3.7	2

4.6.2 Health Vulnerability Within Districts

A sub-district-level analysis was conducted to compare vulnerable areas within a district. Gangnam and Gwanak were selected to represent the deprived and the well-off districts of Seoul.

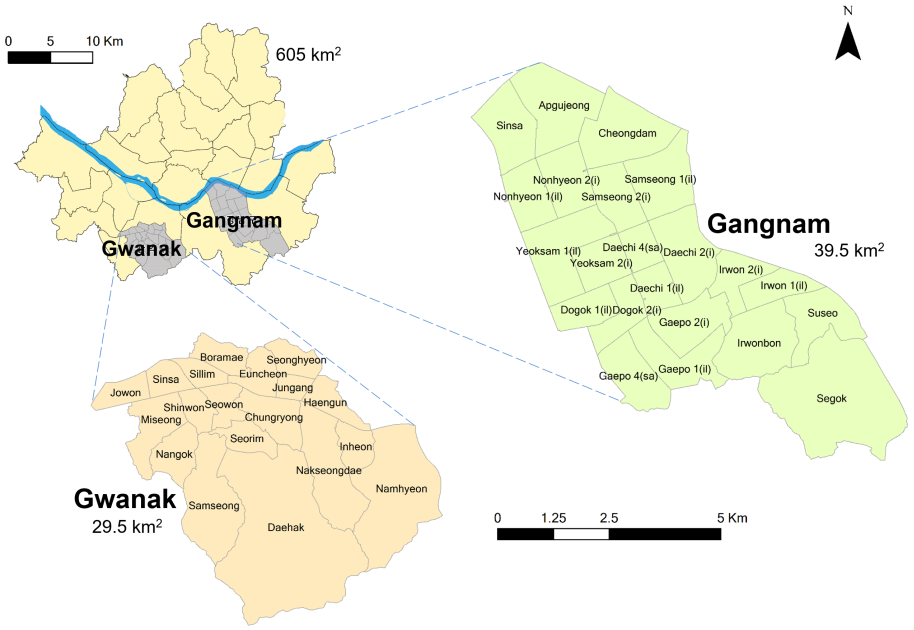


Fig. 4.24 An overview of the study area: Gwanak (bottom) and Gangnam (right)

Gangnam

In Gangnam, geographical variations of health risks were seen within the district boundary, but the top ranked sub-districts of health risk changed dramatically by scenarios.

In the BAU scenario (Figure 4.25 A, B, and C), the fluctuation ranged from 3% to 38% in the AC100 scenario, which was due to the constant repetition of health degradation and recovery of agents whose home or work location is close to a high density of road networks. However, the curves of most sub-districts were flattened in AC150 and AC200 because both the individual and the regional resilience aided residents whose health were under 100. Although Suseo and Segok had reduced health risk by 20% in AC200 compared to that in AC100, both areas still were at the top of the table. Segok had a small population (<200 agents, and the next smallest population

was 695 persons in Samseong1) which significantly influenced the outcome compared to large population districts. Suseo had a demographically young age distribution averaged around the 15-19 and 20-24 age groups: the health loss parameters for these groups were set high.

In the INC scenarios (Figure 4.25 D, E, and F), the temporal patterns of sub-districts were comparable in each resilience scenario, and health risk in most sub-districts did not change in higher resilience scenarios but had slowed the critical transitions to a contrasting rise. The sub-districts including Suseo, Apgujeong, Shinsa, and Cheongdam, maintained the figures at 30-40%, while a few areas such as Samseong1 and Dogok2 managed to have reduced the hazardous rates by 10%. This implies that even though residents of Gangnam had a greater chance of recovery in Seoul, the district still would not avoid the possibility of having a very considerable number of population members at risk as a result of pollution exposure.

In the DEC scenario (Figure 4.25 G, H, and I), all districts have relatively low risk, although again for AC100 Suseo and Apgujeong1 are the most affected areas, but for AC150 and 200 Suseo and Segok stand out.

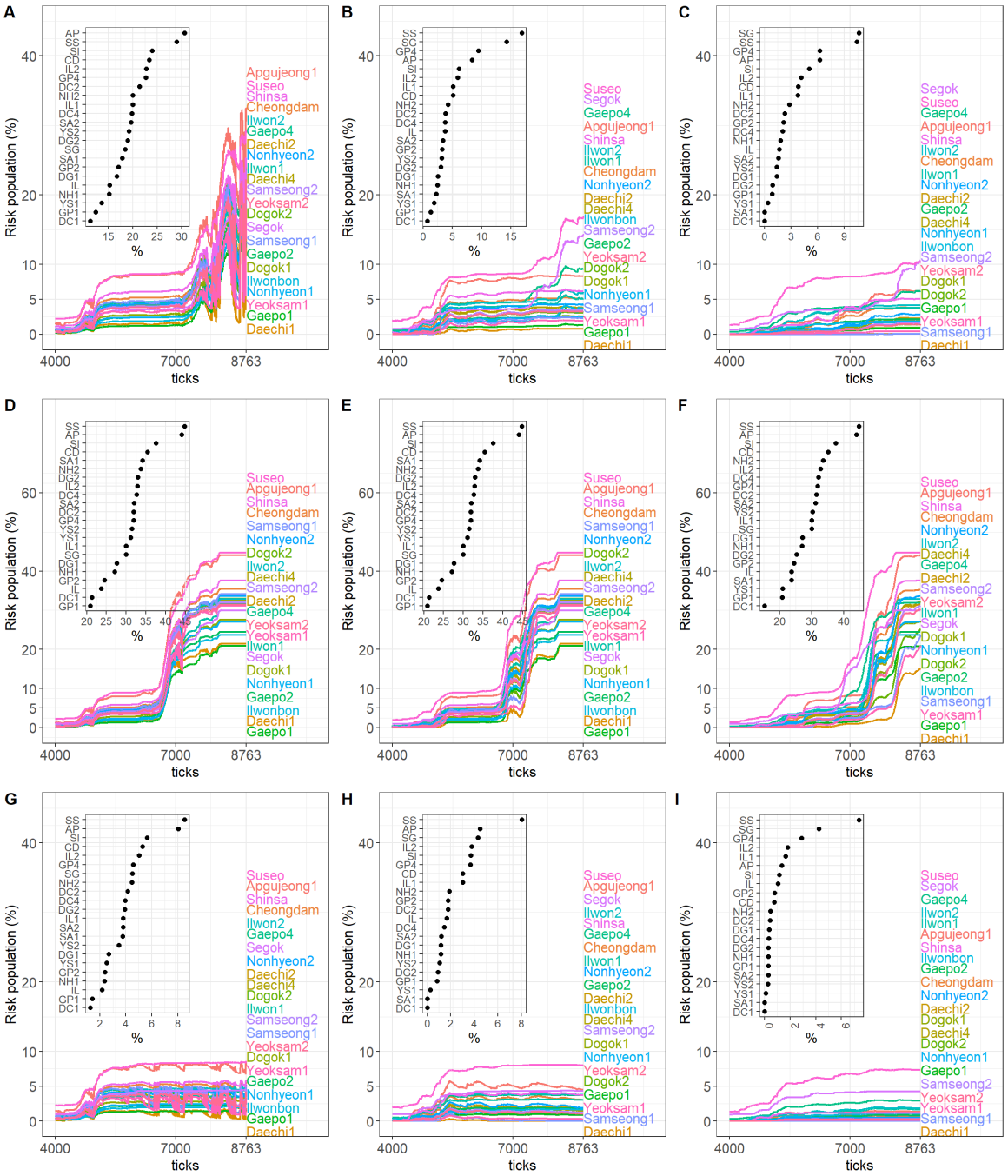


Fig. 4.25 A sub-district result of the risk population in Gangnam. The BAU, INC, and DEC scenarios are aligned horizontally (i.e. ABC, DEF, and GHI), and AC100, AC150, and AC200 are aligned vertically

Gwanak

Gwanak also showed a large variation of health risks within districts with effective resilience, but the difference from Gangnam was not only the percentages of the sub-districts but also that higher resilience was not effective in INC and DEC scenarios.

In the BAU scenarios (see Figure 4.26 A, B, and C), health risks of the sub-districts hardly fluctuated throughout the simulation, however, the appalling increase after 7500 ticks that occurred in AC100 was flattened by half in AC200. The elderly and the youth were again the first groups to have health risks; however, Cheongnim, Shinsa, and Haengun managed to reduce the risk rate closer to 0% when higher resilience was applied.

In the INC scenarios (see Figure 4.26 D, E, and F), the vulnerable groups, along with a lower education level became unwell in the first phase (around 5000 ticks), then an additional 35% of the population became unwell in the second and third phase (6000 and 7500 ticks). Although higher resilience scenarios did not prevent the deterioration of their health, it did help slow the surge.

All sub-districts experienced a small rise of risk in the DEC scenarios (see Figure 4.26 G, H, and I). The highest sub-district was Sunghyun with 7.7% of at-risk population in AC100, but it decreased to 4% in AC200. Compared to Gangnam, oscillations of health risk did not happen in Gwanak due to the weak support of regional resilience, which means that the vulnerable had less opportunities to recover.

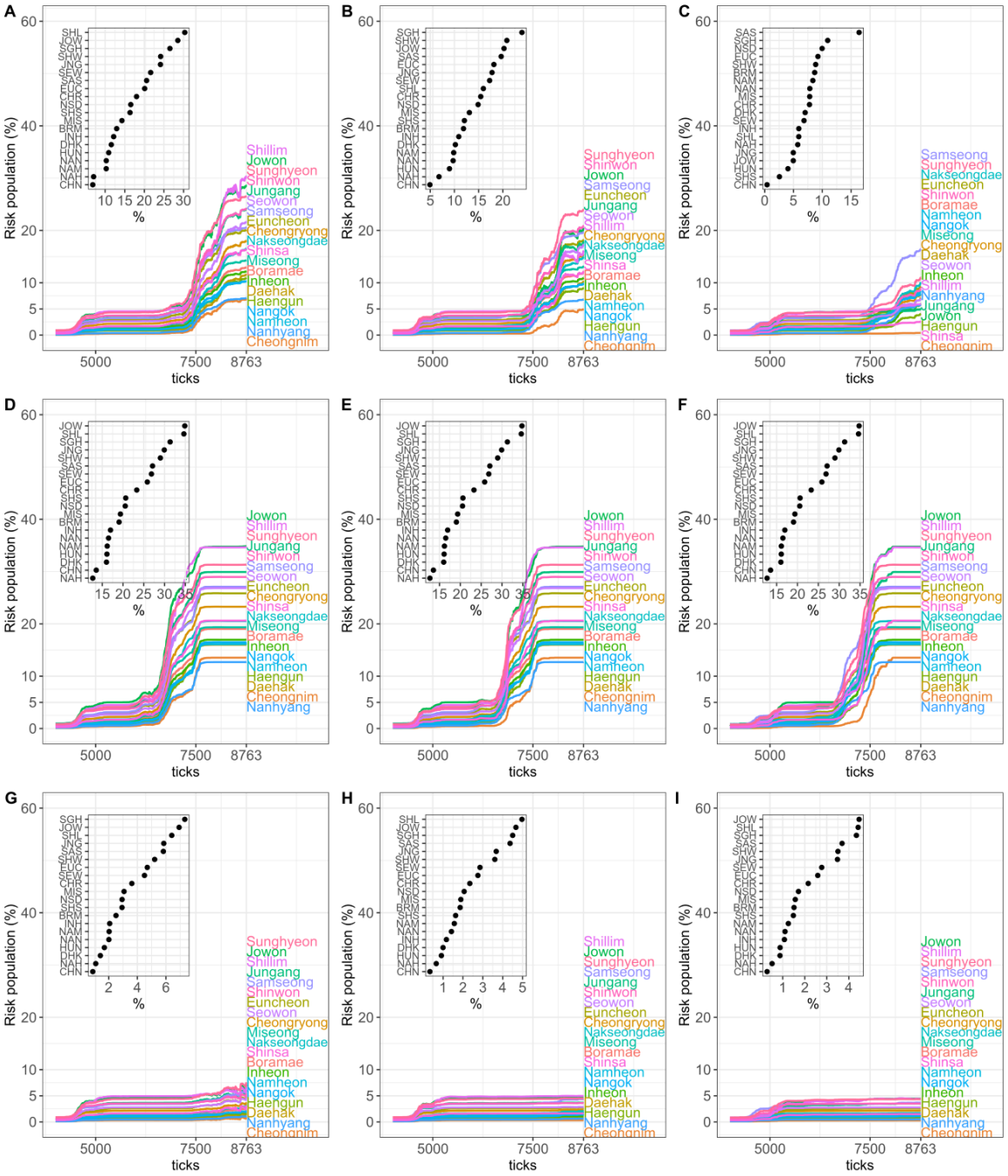


Fig. 4.26 A sub-district result of the risk population in Gwanak. The BAU, INC, and DEC scenarios are aligned horizontally (i.e. ABC, DEF, and GHI), and AC100, AC150, and AC200 are aligned vertically

4.7 Discussion

This chapter investigated how exposure level and socioeconomic resilience affect the population's health in Seoul with an agent-based approach. The study used 5% of the entire population of Seoul, applied OD matrices for each district based on purpose (i.e. commute, school, shopping), assumed that all agents were initially healthy and that a tick was equivalent to half a day. To explore possible future health outcomes from pollution change and regional resilience, the model applied three exposure scenarios (i.e. BAU, INC, and DEC) and three resilience scenarios (i.e. AC100, AC150, and AC200). Simulations were distinctive across all 25 districts.

Prior to producing the scenario outcomes, the study analysed the sensitivity of health loss and additional pollution effect on road patches assuming people's home and work location are nearer to roads tended to trigger the risk. Pollution-on-road had three parameters of 1, 1.5 and 2, in which one was equivalent to the background level, and the result showed that while the health risk of Gangnam resulted in 5%, 12%, and 25% for 0.004, the health risk of same parameter resulted in 4%, 16%, and 19% in Gwanak. This implies that health risks can be sensitive by districts and the parameter values. Not all of the districts were tested, but this study presumes that districts that have denser road networks (e.g. Yeongdeungpo, Guro, and Mapo) may have more sensitivity in their health risk outcomes. Health loss was more sensitive than the road effects because the equation exponentially degraded the health level. Since the nominal health status of 100 was the threshold of 'at-risk', the behaviour of the parameter was sensitive by the unit of 0.001. Gangnam was seen to experience more frequent temporal oscillations than Gwanak partly because the regional resilience was higher (i.e. land price was used as a proxy for better/worse health support), and because Gwanak residents tend to commute to and from roads with added pollution that triggered the aggravated outcome.

Hospital admission was used to calibrate the at-risk population by age groups. Imported from a 1% sample of a nationwide health insurance database named HIRA, this study selected pulmonary diseases to calibrate the simulated results of the default scenario (BAU×AC100), except for the common cold (J00). The study confirmed that most age groups had <200 person difference, which is meaningful in a sense that the model is proven to be possible to support respiratory symptoms and hospitalisation for PM_{10} exposure (Chen et al., 2007; Guarnieri and Balmes, 2014; Halonen et al.,

2016). The agent-based calibration model established in this chapter can help future air pollution and health studies, especially bridging the gaps between the cumulative exposure to (extreme) pollution episodes and health outcomes. The microscopic simulation will enable the trajectories, time of exposure, and the beginning of an onset, and how the emerging patterns differ by demographic groups.

This chapter initially compared 25 districts to investigate how the exposure level and socioeconomic resilience affect the population's health. As a result, the vulnerability in all districts experienced an early onset at 4000 ticks then skyrocketed after 7000 ticks. This was strongly related to when the demographic group entered the unwell (i.e. health under 100) phase as well as how much health recovery is supported from the district. Furthermore, the vulnerability rate of the elderly and the lower-educated increased by over 10% in the BAU scenarios, over 30% in the INC scenarios, and over 5% in the DEC scenarios regardless of resilience scenarios. These findings extend those of Jerrett et al. (2001) O'Neill et al. (2003), and Moreno-Jiménez et al. (2016), showing that disparities in health outcomes are likely to depend on socioeconomic status (SES), especially when the group is exposed over a long period geographical factors become important (Brook et al., 2004; Jerrett et al., 2001). It is also important to take into account the unchanging figures of the vulnerable population such as those of the elderly because these people, in fact, may not have good access to adequate health care during a pollution episode.

With a relatively severe 3% per annum increase rate in pollution loading, this chapter found that the affected population suddenly rises after 7000 ticks, i.e. 10 years. If pollution exposure is actually linear with the rate of change of health, then this suggests that health services may be suddenly overloaded if pollution levels are allowed to continue rising. However, a decrease of pollution levels at AC100 in some districts, such as Gangnam, gave rise to larger oscillations in the population measured to be at risk. This is disturbing, as it suggests there may be some regimes in which stress on health services is increased by lower pollution levels, as the level of variation in the population presenting with symptoms might be difficult to cope with. On the other hand, the mean level of population in districts where residential areas are near to commercial and roads remained consistently high.

A sub-district analysis identified vulnerable areas within Gangnam and Gwanak: wealthy versus deprived. Gwanak sub-districts showed a uniform pattern in health risk outcomes: high risk areas remain high in all scenarios and the rank never changes.

Gangnam, by contrast, had a large level of spatial and temporal variability by resilience scenarios. The high variation of land prices, population, and road networks across the district may have determined the outcome. Overall, the scenario results indicate that the health risk in districts such as Gangnam might be more easily controlled through a policy application of pollution alerts and clinical care, but Gwanak is likely to be in more urgent need of attention because there were less oscillations happening. This means that the district itself has less resilience to cope after being badly exposed to air pollution. However, the range of sub-districts in the outcome implies that outcomes measured by aggregating over populations can be highly scale-dependent, thus caution is needed when finding an appropriate demographic scale.

This study is the first, to the author's knowledge, to examine health vulnerability with pollution for individuals, taking into account the variability in adaptive capacity to be expected from heterogeneity in social class and its spatial realisation. In addition, significant improvements were made over the Urban Suite-Pollution model (Felsen and Wilensky, 2007), particularly in the application of road effects, OD matrices, landcover, and hospital admissions data. However, this study only applied a population proxy to the model, with a randomly selected value of a single pollution site for every location. The latter will have inaccurate pollution fields: actual pollution fields are generated preferentially along major highways and vary temporally, for example with the rush hour, and by transport mode, for example, bus versus car. Nevertheless, the settings and findings were clear to stand as an *'illustrative model'*: not to mimic the reality of people's health vulnerability and locational pollution level in Seoul but to understand how health outcomes by districts and demographic groups can vary when pollution exceeds the legal threshold (Edmonds et al., 2019).

Furthermore, the representation of the health response to accumulated exposure focused on the need to assess the exposure 'dynamics', which involves spatial and temporal variations based on personal mobility and time-activity patterns. In the model, this chapter assumed the exposure level to be a result of consistent interactions between the air pollution patches and the agents' trajectories. This has not been covered in prior studies based on fixed-site monitors, which produce relatively coarse 'static' measurements. However, the actual translation of exposure into morbidity is an area that is very complex: little is really known about how short, high-intensity exposure features in importance relative to lower-level, long-term pollution loads, for example. The detail of the results should, therefore, be treated with a great deal of

circumspection; however, the model does illustrate that the combination of spatial variation in pollution loading, population movement and health recovery may lead to complex and possibly counter-intuitive patterns in the risk of disease over time and over space.

Future work should therefore include not just a more carefully developed synthetic population and more sophisticated dose and response models for different types of pollutants – for instance David and Don (2012) – but also conduct a traffic simulation to generate the pollution field dynamically, constrained by the emissions data. In this way, the feedback between the traffic flow that generates pollution and the exposure to that pollution can be linked to the behavioural characteristics of the agents and to policy changes aimed at mitigating emissions. For example, the creation of a traffic-free zone in a central district might clear the air locally but lead to changes in traffic flow that increase the exposure in other locations. This kind of effect would not be reproducible with a simple statistical micro-simulation that relates past observed pollution and traffic flow to exposure. The author hopes that the interaction between life histories, social circumstances, the choice of residential location, the transport mode, and policy for city development can then be integrated to investigate how pollution loads might be reduced in a socially just and spatially sensitive manner.

Chapter 5

Exposure to Traffic-related Air Pollution in Central Seoul using an agent-based framework

Overview

This chapter constructs a traffic simulation for Central Seoul to understand the coupled problems of emissions, behaviour, and the estimated exposure to PM_{10} for groups of drivers and subway commuters. Compared to the previous chapter, this study adds the movements of individual agents as well as the dynamics of generation, dispersion and dilution of particles that can harm human health, but maintain the concept of building an 'illustrative model'. Non-exhaust emissions from brake and tyre wear are used since these are unavoidable even for vehicles with low or zero exhaust emissions.

The results show that significant extra particulates relative to the background exist along roadways where non-exhaust emissions contributed 40% of the total roadside PM_{10}. This led to the acute health risk of resident drivers on high air pollution days, where 88% of them were at risk in late March 2018; however, these spikes rarely happened. By contrast, the health risk to subway commuters peaked at a maximum of 30% but showed more frequent oscillations during air pollution episodes due to a wide range of commute times. However, these estimates of health effects depend strongly on how impact and recovery from exposure are parameterised. A vehicle banning scenario dramatically reduced the risk to resident drivers by 30% of PM_{10}, but it was

not effective for subway commuters, as the group generally walked through background areas rather than along major roadways. Changes in pedestrian behaviour did not seem to result in significant mitigation of exposure.

Using an agent-based traffic simulator in a health context can give deep insights of how exposure and health outcomes can depend on the time of exposure as well as the mode of transport, and suggest what might be the consequences of behavioural or policy changes, unlike aggregated outcomes from low spatial density observations.

Keywords: traffic simulation, non-exhaust emissions, exposure, health effect, Central Seoul

5.1 Introduction

Traffic-related air pollution (TRAP) has long been associated with adverse health outcomes. This has been evident by various case studies that used NO_x, O_3, PM_{10}, and $PM_{2.5}$ collected from stations, surveys, or pollution sensors (Altuğ et al., 2020; Atkinson et al., 2016; Min et al., 2020; Samoli et al., 2016; Shou et al., 2019; Stafoggia et al., 2013). While many studies have concentrated on gaseous pollutants and particulates from exhaust to investigate health outcomes, recent studies have also insisted that non-exhaust emissions (NEE) are a threatening source to human health (Air Quality Expert Group, 2019). The pollutants released from brake and tyre wear can easily enter the human respiratory tract, deposit on the lung walls, and eventually impair the lung airways (Altuğ et al., 2020; Petkus et al., 2019).

Recent exposure studies have interpreted exposure to TRAP based on a daily aggregation of exhaust emissions because the gaseous pollutants show significant contrasts between the daytime and nighttime, which is likely to be contributed to by traffic (Atkinson et al., 2016; Min et al., 2020). These gaseous pollutants have shown a convincing outcome in both short- and long-term exposure assessments, but the association between particulates (PM_{10}/$PM_{2.5}$) and adverse health outcomes was hardly noticeable in short-term research (Atkinson et al., 2016). Even studies that attempted to associate TRAP with hospital data, focussing on annual mean PM_{10} only gave a generic association with mortality or hospital admission; however, relying on annual means may not fully allow an understanding of whether a disorder was gained from acute exposure to short term pollution extremes or to a constantly high (but

below a warning level) of pollution (Briggs et al., 2008; Gulliver and Briggs, 2007; Shou et al., 2019).

Since previous studies have preferred to import pollution data from monitoring stations, exposure estimates might be inaccurate with respect to where people are actually located. Although spatial interpolation predicts a value of an unknown location, pollution data collected from 'spatially sparse and temporarily rich' stations normally give spatial errors (Hoek et al., 2008; Min et al., 2020; Xu et al., 2019). The position of the sensor (e.g. near the road, high up on a telegraph pole) can mislead our interpretation of the pollution effects at human height. Moreover, the measurement toolkits might show a discrepancy for where the person is located (Lee and Park, 2019). For instance, an instrument called a β-absorption monitor is installed in most of the stations because all the measurement processes are automated, but it requires at least an hour to capture and export the data, which is not useful for real-time measurement.

NEE studies have discovered that traffic queuing near junctions and stop-and-go behaviours have largely contributed to the local atmospheric particulate load (EMEP/EEA, 2019; TfL, 2018). Panko et al. (2013) discovered that less than 1µg/m^3 of road PM_{10} was contributed by NEE in cities of Japan, France, and the US; however, the airborne samples may not have captured the sources contributed from the surface level which contains most of the particulates. Recent studies are still finding evidence from lab work (Alves et al., 2020; Kwak et al., 2013; Panko et al., 2013; Perricone et al., 2018), tunnels (Lawrence et al., 2013), and selected roads (Ferm and Sjöberg, 2015). These studies claim that the metallic pollutants Zn, Cu, and Pb can contribute to the local atmosphere and possibly harm human lungs, which is credible, but will depend largely on the volume of vehicles which in turn increase the frequency of the friction of brake discs from stop-and-go behaviour.

To link the challenges between disciplines, associating the exposure to NEE over a longer period with the dynamics of movement patterns is required. Agent-based modelling (ABM) is one method that can simulate the interplay between air pollution and the effects on a population's health (Tracy et al., 2018). Instead of assuming pollution data from a distant location applies to the current location of an individual, this tool simulates the movement of heterogeneous vehicles and individuals and measures the exposure level based on the path on which the agent is situated and the estimated local pollution value (Crooks et al., 2019; Gurram et al., 2019; Wilensky and Rand, 2015). This can improve on the problem of ecological fallacy in which aggregate

population-level effects are assumed to apply to individuals, and also account for feedback effects between changes in behaviour and the accompanying generation of pollution load. A few traffic simulation studies have investigated human exposure to NO_x: Gurram et al. (2019) used MATSIM to model a single day (since the model is computationally expensive), but the duration was too short to consider the exposure of every individual by their activity patterns relative to longer timescale variability in pollution load. Yang et al. (2018) applied ABM to examine how participant's exposure to NO_2 can vary by the selected modes of transport. Rather than comparing exposure by transport modes and their routes taken, the interpretation of the results was too reliant on seasonal effects, which can also be seen in the monitoring station data (Yang et al., 2018). See Section 5.2.2 for more detail on these models.

This chapter examines the exposure and possible health effects of NEE on commuter's health based on a traffic simulation approach. The specific questions are as follows:

- What is the difference in health effects between walking commuters and vehicle commuters?
- How to model the spatial distribution of NEE generated by traffic flow?
- How sensitive were the parameters for vehicle emissions and the response of humans?
- How did air quality improve as a result of the simulation of policy scenarios, and what were the characteristics of any improvements?

Given the limited resources available to mimic the agents' attributes and their behavioural patterns, this study attempts to build an *in silico* agent-based traffic model to illustrate how likely are vehicles to emit and disperse PM^{10} to their local atmosphere, and to what extent the agents are affected from vehicle emission and background air quality.

5.2 Literature Review

5.2.1 Non-Exhaust Road Emissions and Health

The book *Non-exhaust emissions: an urban air quality problem for public health* compiled recent scientific findings that addressed non-exhaust particles that are generated in forms of metallic, rubber, carbon black, and other organic substances by combustion, wear[1], road abrasion[2], and dust resuspension[3], all of which are equally as potentially catastrophic as exhaust particles (Amato, 2018). Although the regulation for NEE has not yet been made, both the UK's Air Quality Expert Group (2019) and the European Environmental Agency (2019) have noted the potential severity of NEE for human health and are calling for more evidence.

NEE can be affected by traffic queues, driving behaviour, and weather. In traffic congestion, the stop-and-go patterns of the traffic generate more wear on brake pads and discs that adds to surges of ambient particulates during rush hours (Air Quality Expert Group, 2019). A correlation between traffic weight and PM_{10} has also been found in Cambridgeshire (Cambridge Insight, 2017). Brake wear emissions are also spatially heterogeneous because the vehicles would be expected to slow down when reaching a junction or going downhill (Air Quality Expert Group, 2019; Smit et al., 2010).

Second, aggressive drivers are apt to accelerate and slow down more which eventually results in pollution with particles from tyres, brake discs, and linings. Transport for London has mentioned that the Central Business District (CBD) of London is more polluted than the outer areas, but since the average speed of the vehicles does not exceed 20mph, PM_{10} can possibly vary according to driving behaviour (TfL, 2018).

[1]Tyre wear generates particles when the rubber erodes the road surface. The particle size varies, but a great proportion is below 10 micrometres in diameter, which can be apt to contributing non-exhaust particles to the atmosphere (Amato, 2018). Brake wear occurs when the brake discs and linings are heated every time the driver steps on the brake pedal (Amato, 2018). Brake wear is highly correlated first to the vehicle's velocity because more friction is needed to slow the car down, and second to driving habits in which fierce acceleration and sudden braking can also release particles into the atmosphere.

[2]Road surface wear is a form of friction between the road surface and tyre surface, and the particles are more likely to abrade when the road condition is bad or where heavy trucks damage the road surface.

[3]Resuspended dust comes from various sources including soils, small metals, dirt, and even microplastics and other objects from vehicles.

This is also why EEA's brake wear test shows that PM_{10} emissions are high because drivers are more likely to stop-and-go when the speed is below 40 (EMEP/EEA, 2019).

Finally, NEE is found to be dependent on seasonal effects. In northern Europe, where the roads are frequently frozen, studded tyres generate harmful particles as the metal strikes the road surface, which leads to an increase in dust resuspension with the sand (rock salt) previously scattered (Air Quality Expert Group, 2019; Amato et al., 2014; Weinbruch et al., 2014). These are examples of road abrasion. While other parts of the world do not use studded tyres as much as northern Europe, a few countries are experiencing higher road particulates from resuspension in drier seasons. South Korea is a good example of this case because the country has a humid summer with monsoon seasons as well as very cold and dry winter. Seoul has been using high-quality sensors underneath vehicles to monitor roadside pollution and sends water-spraying trucks to wash away particles if $PM_{2.5}$ exceeds $200\mu g/m^3$. This has been useful to some extent, but with limited monitoring vehicles and cleaning trucks and a high demand for water-spraying in winter seasons, dust control has been a continuous problem to the city. Moreover, it has been pointed out that the water-spraying effect only lasts for a couple of hours.

Despite efforts to reduce particulate matter from vehicle exhaust, little improvement has been made for non-exhaust particles from tyre and brake wear and dust resuspension. Ferm and Sjöberg (2015) applied an equation of particle emissions based on roadside PM_{10} and NO_x in two Swedish cities to associate ambient PM_{10} (both road and background) with the volume of vehicles; however, they found it had little relevance. The problem for this study was that the modelled PM_{10} emissions were highly dependent on NO_x because the observed NO_x, which assumed to have been generated by traffic, had an hourly level variability that is not seen in PM_{10} observations. Also, the experiment was site-specific and conducted on a 500m road. This can either over or underestimate the effects of the road especially when the study domain is small. Panko et al. (2013) discovered a small contribution from non-exhaust particles in French, US, and Japanese cities ($< 0.7\mu g/m^3$ of PM_{10}). Despite being one of the early works on NEE, the selection of locations was biased toward parks, residential areas, or places of worship, which underestimated the effect of NEE. This is because these areas are mostly far from the road and NEE are known to be affected by 150m (WHO, 2013). Electric vehicles are increasingly seen as an alternative to petrol vehicles, but findings

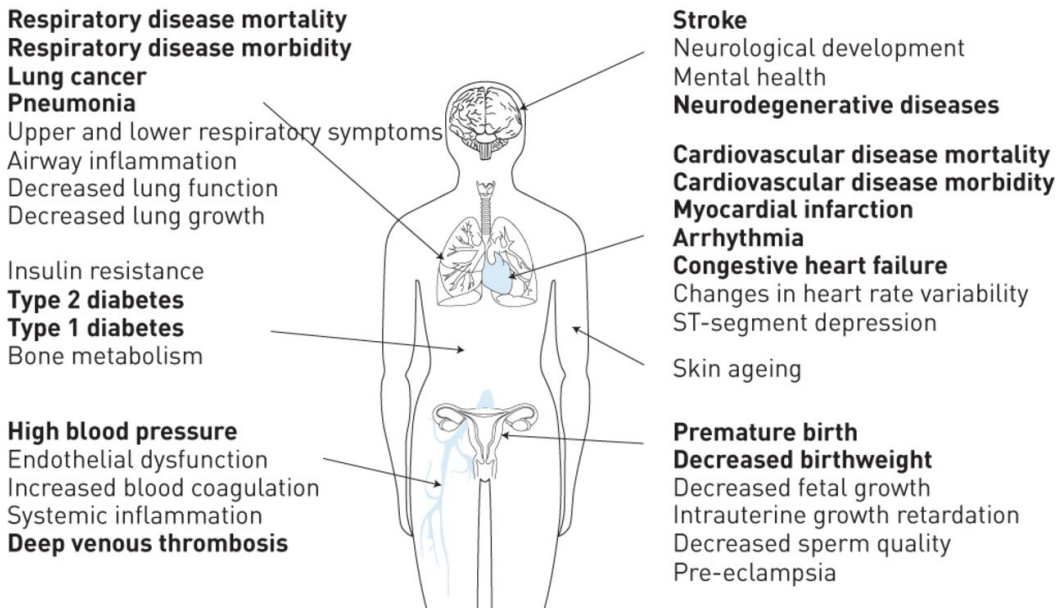


Fig. 5.1 Overview of diseases, conditions, and biomarkers attributed to ambient air pollution (G.D. et al., 2017)

demonstrated that they emitted more non-exhaust particles due to their heavier weight (Timmers and Achten, 2018).

From a health perspective, Amato (2018) reviewed that, despite the limited extant literature, NEE (or road particles, as written) were associated with hospitalisation risk, where the excess risks rose by 4.5% for every extra 1.71µg/m^3 and 2.1% for cardiovascular admission in New Haven, Connecticut, USA; this was also found to be true for Hong Kong. The point here is that roadside PMs are generated by brake and tyre wear that can cause oxidative stress to pedestrians walking along the kerbside (Atkinson et al., 2016; Borm et al., 2007). Oxidative stress is one possible effect of the penetration of toxic particles into the human body. Once the fine dust is inside one's body, it can reduce the size of airways, which eventually impairs lung capacity (Khajeh-Hosseini-Dalasm and Longest, 2015). If the dust is too fine to be trapped in the lungs, then it will flow into the bloodstream, which can cause cardiovascular problems. Other than stress, long-term exposure without treatment can lead to diseases such as depression (Altuğ et al., 2020; Petkus et al., 2019), dementia (Shou et al., 2019), or other diseases shown in Figure 5.1.

Despite the findings of epidemiological studies that associated road dust with oxidative potential or other diseases, there is still an ongoing debate about whether

the disease is solely attributed to NEE or a mixture of gaseous emissions, or whether the person who had an underlying disease might perhaps have become unwell after being exposed to polluted air. Perricone et al. (2018) reported that brake wear emissions resulted in 8–27% of the total traffic-related PM_{10} emissions, and that airborne particles generated from the friction of brake discs can lead to adverse health effects accordingly. With the start of the REBRAKE project, they introduced a concept paper to reduce PM_{10} emissions for car brakes by 50%, leading to a 4–14% reduction in PM_{10}. Other studies are still conducting experiments on the impact of NEE in a restricted environment, such as a laboratory (Kwak et al., 2013) or in tunnels (Lawrence et al., 2013). However, in a laboratory environment, it is easy to measure the emissions due to brake or tyre wear, but the uncertainty may vary depending on the actual road surface condition; more importantly, the tests can be time-consuming and may not well represent the actual driving conditions where weather conditions and other compounding factors affect emissions.

In summary, NEE has recently been highlighted as a new threat in vehicle emission studies because particles are well known to cause pulmonary and cardiovascular disorders, and the chemical substances generated by brakes and tyres can trigger adverse health outcomes. Despite the growing importance, previous studies had 1) difficulty capturing the evidence for NEE in urban settings, both indoors and outdoors; 2) no evidence of considering the dynamics of movement; and 3) less variation in geographical (climate) locations, which can be important in road particle distribution.

5.2.2 Agent-Based Traffic Simulations and Pollution Exposure

Traffic simulations have been used to estimate population exposure, using the dynamics of the movements of heterogeneous agents. Of 30 studies of traffic simulation screened, two studies were identified to have quantified population exposure to air pollution based on traffic simulation (Gurram et al., 2019; Yang et al., 2018).

Yang et al. (2018) examined the cumulative exposure to environmental stressors, namely NO_2 and temperature, using ABM in Hamburg. With eight individuals having different commuting preferences between car, bike, and public transport, the selection of transport was made based on the preference of choice between transport modes. The exposure to NO_2 was compared between summer and winter months on a daily

basis. It appeared that the overall exposure to NO_2 during the winter period was 10-12 times higher than during summer, although the patterns between individuals varied by the means of transport used; temperature or heat waves, which were assumed to cause an increase of stress, were not so high as to adversely impact exposure levels because the temperature of Hamburg did not exceed 34°C frequently.

Although ABM was used as a tool to choose the probability of transport mode, the probability of preferred transport was heavily biased towards one mode, i.e. the agents who either used personal vehicles or bicycles continued to use the same transport mode throughout the period. This, in turn, did not see many changes in exposure levels, particularly in winter when NO_2 was noticeably high. Besides, although the assumption that varied exposure levels according to transport modes was realistic, the interpretation oversimplified the granularity and dynamics that ABM can offer. For example, "the agents will commute to the city centre where high temperature and NO_2 level are expected" does not fully explain which path or transport type affected the exposure level. As another example, with "the average exposure to NO_2 is greater in winter than in summer, while the differences vary by behaviour" also does not account for the effect of vehicle emissions or pollution in the background but only mentions the changes in an abstract way.

Gurram et al. (2019) incorporated transportation, emissions, and exposure modelling to accurately simulate population exposure in Tampa, Florida, USA (see the framework in Figure 5.2). First, the activity-travel patterns of the entire population were generated by DaySim, where the demand varied by demographic status (e.g. race, income) and activities (e.g. school, work, other activities, modes of travel between activities). Since DaySim only calculated the space and time of activities, MATSim (Multi-Agent Transport Simulation) supported the movement of an individual's daily schedule. MOVES (MOtor Vehicle Emission Simulator) and R-LINE (Line Source Dispersion Model) toolkits were employed to emit and disperse NO_x during the journey. Exposure was measured by the ambient concentration and the activity record on a daily basis, where the activity record included location, time of arrival, and duration at a fixed location.

Results showed that the mean exposure was lower than their earlier study (Gurram et al., 2015), whereas the maximum exposure was remarkably higher. Unlike the CALPUFF dispersion model exploited in their previous research that considered the details of the complexity of terrains and meteorological pre-processors, the agent-based

model was able to capture the instantaneous rise of NO_x with a high-resolution activity-based exposure. This study revealed that the exposure levels varied according to socioeconomic status: "Higher exposure was discovered for the lower-to-middle income households, people with colour, and people of working ages (19-65).", and for those who lived in the urban zone with longer travel times. Overall, integrating each of the four phases of different simulation platforms is ground-breaking for simulation-health research, particularly for capturing instantaneous exposure (hourly periods) than that of aggregate exposure, which hasn't been investigated in short-term exposure studies.

Despite the novel construction of a traffic simulation coupled with the activity and emission models, some points remain problematic. First, the 24-hour simulation was too short to capture the exposure based on people's activity but likely depended on one's socioeconomic status and place of residency. As the authors found no noticeable rise of NO_x over the whole simulation, the only possibility of pollution exposure was likely to occur in the city centre or areas adjacent to major roads. These areas were mostly of a lower- to the middle-income background or non-white. Second, the simulation only considered the activities of residents who use passenger cars; however, it did not include the travel of non-resident vehicles or residents who chose other modes of transport (transit, biking, walking). From the evidence from Seoul, more than 6 million trips were made by vehicles that commuted to and from Seoul, which comprises a third of the total traffic of 2010. Neglecting the number of pollutants generated from this group might have resulted in much of the local emissions that contributed a significant amount to the ambient pollution being omitted. Inclusion of other transport modes may sound plausible, but this will entirely depend on the simulation capacity.

In summary, it turns out that study of the linkage between vehicle emissions and human exposure is still young, but these two studies have opened up a new discipline that blends the distinctive topics to better understand the reality of human exposure. Indeed, the preliminary study from Hamburg needs greater sophistication in its simulation design and implementation to fill the gaps that individual mobility studies with GPS cannot answer, e.g. an experiment with hundreds of heterogeneous entities and their journeys despite a limited trip-related data available, which route was most polluted, how do faster commute times with more chances of high NO_x exposure compare with the opposite case, and what message does this study deliver in terms of the association between choosing a transport mode and being exposed to environmental stressors (Mastio et al., 2018). The second study was impressive regarding the massive and

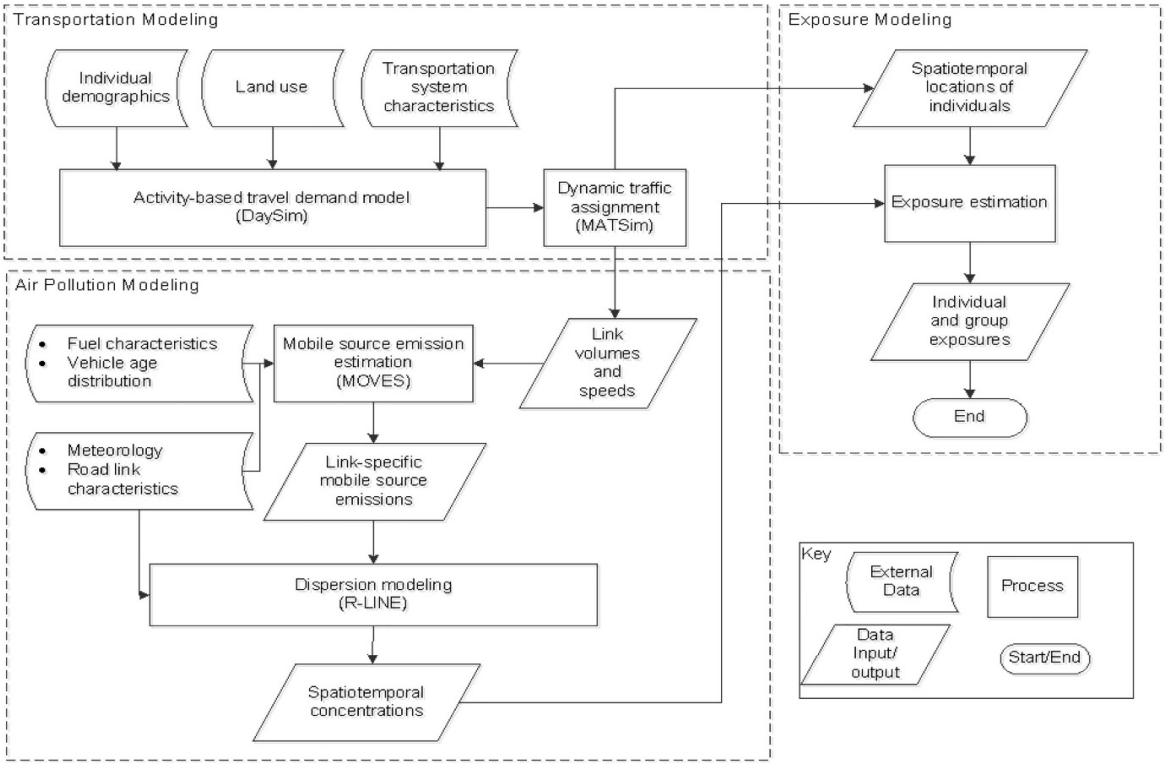


Fig. 5.2 The integrated agent-based transportation and air quality modelling framework for population exposure estimation (image imported from (Gurram et al., 2019)

sophisticated construction of the individual activity patterns of a synthetic population, simulating their journey with a traffic simulator and calculating exposure based on the time spent in each microenvironment – something that nobody has ever attempted. However, the results that low-income and non-white people were more exposed to NO_x tended to be too deterministic from a 24-hour simulation; since NO_x is highly related to traffic, excluding the non-resident traffic may have underestimated the pollution levels.

5.3 Data Collection and Exploration

5.3.1 Study Area

The study area for this chapter covers most parts of Jung and Jongno district (16.6km^2) which is the CBD of Seoul. This CBD zone also refers to as the "Sadaemoon" area, which means the four gates from the previous Josun Dynasty. There are eight sub-districts[4] in the Jongno district and seven sub-districts[5] in Jung. This area plays a central role in business, tourism, shopping, and transit to sub-civic centres.

To improve urban air quality, Seoul planned a Green Transport Scheme which set a clean air zone within the four gates of central Seoul. Unlike London's ULEZ (Ultra Low Emission Zone), the government has fined vehicles that do not meet the emission standards since December 2019. Despite the attempt to provide better air quality for citizens and commuters, there are still concerns about whether the reduced traffic has benefited air quality since the commencement of peak hours, or has caused another problem for those drivers whose vehicles do not meet the standard who coercively have to change their mode of transport.

Seoul CBD is was chosen because first, approximately 300,000 people and 1.3 million vehicles commute to the area everyday, second, direct and indirect effects from exhaust and non-exhaust emissions release notorious levels of particulates which can affect residents' and commuters' health, and lastly, hardly any scientific studies have discovered the association between vehicle ban and air quality after the 'Green Transport Scheme' was initiated. More TRAP studies, such as London Low Emission Zone (LEZ) (Beevers, 2016; Beevers et al., 2013; Halonen et al., 2016; Kelly et al., 2011) are needed. This study is hoping to contribute on non-exhaust emission and the instantaneous rise of PM_{10}, but also play an important role to use potential scenarios to envisage traffic control (by fuel types or vehicle size) and the improvement of air quality.

[4]Sajik, Cheongwoonhyoja, Samcheong, Gahoe, Jongno 1-4ga, Jongno 5-6ga, Ewha, and Hyewha
[5]Sogong, Hoehyeon, Myung, Pil, Jangchoong, Gwanghee, and Euljiro

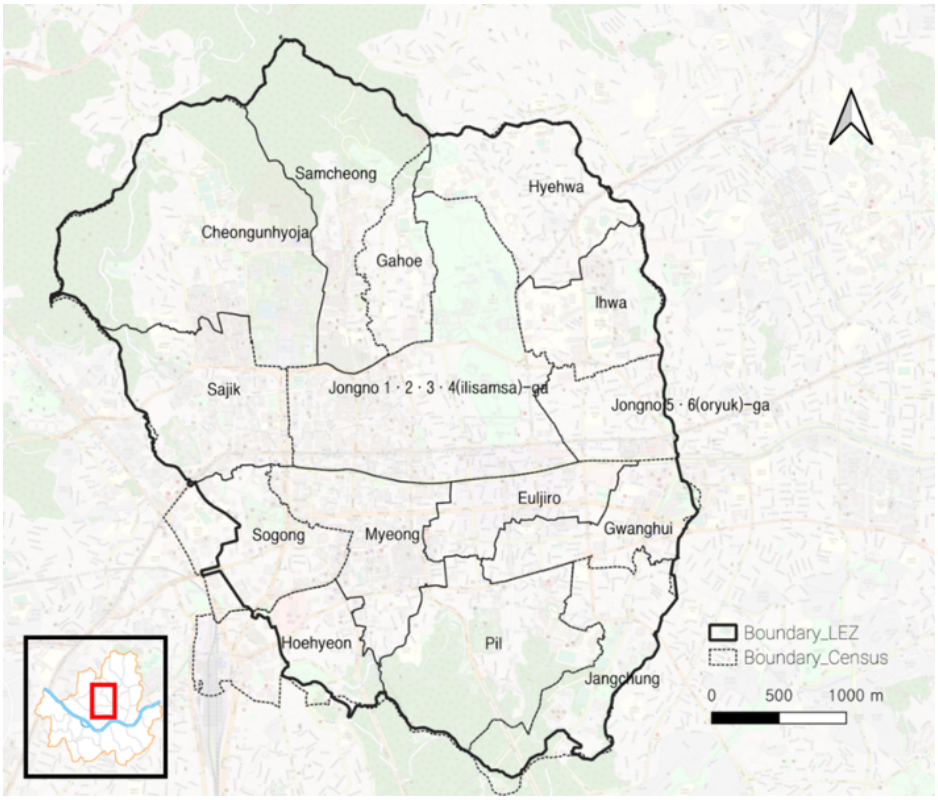


Fig. 5.3 Study boundary – Seoul CBD

5.3.2 Data Collection

Air Pollution

Hourly measured PM_{10} was imported from two urban background stations (Jongno, Jung) and two roadside stations (Jongno, Seoul Station) provided by the National Institute of Environmental Research. Assuming subway commuters travel outside the CBD area, PM_{10} was also collected and averaged from 23 background and 12 roadside stations within the city boundary. The collection period was set between January 1st-March 31st, 2018, considering high concentrations and variations of PM_{10} during the winter period. The study chose the three months for which the data were available in the Seoul Institute repository.

Population

To investigate the dynamics of instantaneous exposure to PM_{10} by different commute patterns, subway commuters and resident drivers were compared. To obtain an adequate population for both groups, this study separated the approaches for data collection. The use of sampled agents was to maintain the computation speed and the health risk rate which did not change to a large extent when the sample size was bigger.

For subway commuters, 1% of the census and the daytime population of Jongno and Jung districts were imported from national statistics. To obtain a closer estimate for subway commuters, this study attempted to subtract the daytime population by the census, then multiply the fraction of subway users. Here, Jongno and Jung had 147,607 persons and 118,632 persons reported on the census, and 392,128 persons and 442,225 persons for the daytime population. This can be summed up to 266,239 persons and 834,353 persons respectively. Daytime population subtracted by the census equals to 568,114 persons. This can be multiplied by the rate of subway commuters (34%) which returns 193,159 persons. Hence, the model imported 1% of the final population of 1932 agents in the simulation. However, it is worth noting that the rate of subway commuters was provided by the Korean Transport Database that surveyed transport modes used for commuting to the destined admin zone during weekdays, which offers counts of transport modes taken from the participant rather than considering the total number of transit taken from a single person. For example, a participant who took a taxi and the subway would have answered both.

For resident vehicles, each sub-district was provided with a sum of vehicles by fuel type that include gasoline, diesel, LPG, Hybrid, and so on (Seoul Statistics Office, 2019). Having excluded the vehicles whose administration zone had recently incorporated with others due to population change, and a final sample of 399 registered vehicles, close to a 5% sample, were imported for simulation.

Traffic Count

Traffic count data was provided by the Transport Operation & Information Service (TOPIS) of the Seoul Metropolitan Government, with the intention to implement the influx of non-resident vehicles into the study area. The statistics were recorded on an hourly basis at 24 observation points that start with A01, A02, and so on (see Figure 5.4). The attributes include dates, hours, flow directions, and road names. To

get accurate counts of data but to avoid duplication with closer stations, 10 stations were imported into the simulation: A02, A03, A04, A07, A08, A09, A10, A12, A20, A24.

Since the model uses 'minutes' as a temporal unit, the hourly traffic counts were decomposed and rounded to the nearest minute. For example, on the 3rd of January 2018 at 8:00-9:00am, there were 3,075 inbound vehicles at A-02 (Sajik road), which was divided by 60, and resulted in 51 vehicles per minute on average that flowed into the study area.

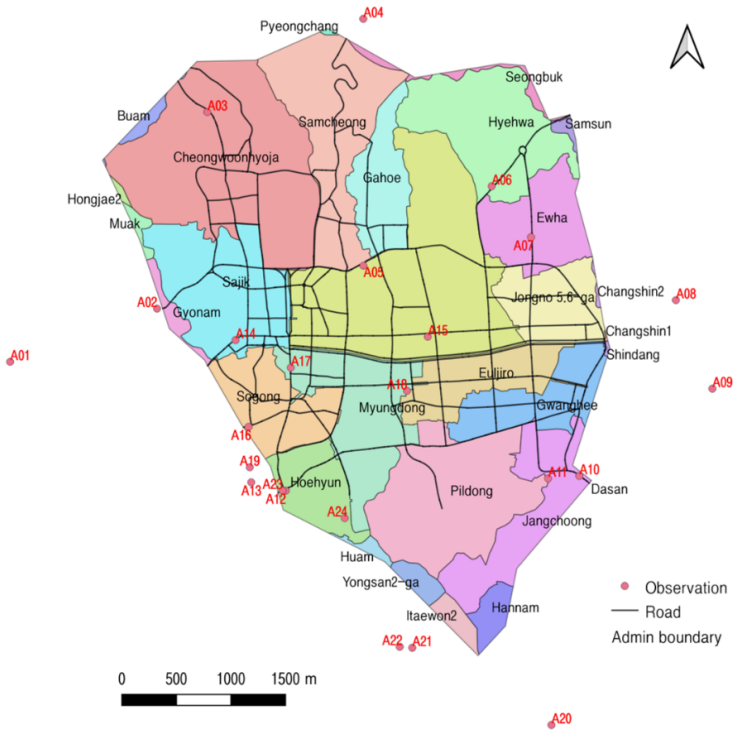


Fig. 5.4 Traffic count monitors around the CBD area from A01-A24

Korean Transport database (KTDB) provides two types of origin-destination (OD) matrices: *mode-based trips and purpose-based trips*. Both matrices present the estimated counts of passengers in a day, e.g. Admin Origin: Sajik, Admin Destination: Jongno, Type of Mode: Subway, Count: 38,541. A mode-based matrix from 2015 was selected over a purpose-based matrix because the study focus is on the journeys by transport mode.

Amongst 15 modes[6], the matrix for passenger cars was selected as proxy data for mobile agents. As this study used a fraction of the population in the model, the raw data was converted to a fractional OD matrix. However, it is noted that the administrative boundaries in 2015 might not always match the boundaries of the daytime population or vehicle registration zones in 2018 due to rapid population change that resulted from urban regeneration and gentrification. The OD matrix for subway commuters was omitted because the destinations at the subdistrict level were already assigned in the setup.

5.3.3 Data Exploration

Air Pollution

The time series of PM_{10} between January-March 2018 was provided from two urban background stations, Jongno and Jung, and two roadside stations, Jongno Road and Seoul Station (see Table 5.1 and Figure 5.5). Overall, PM_{10} at Seoul Station was at least 10µg/m^3 higher than the other three stations on average, while Jongno Road, although installed close to an eight-lane road, had similar pollution levels to the background measurement.

In January, Seoul Station (58.2±33.7µg/m^3) had more PM_{10} than the other stations by 11-14µg/m^3 on average (Jung: 44.6±23.9µg/m^3, Jongno: 44.8±24.3µg/m^3, Jongno Road: 47.3±26µg/m^3). The concentrations in February and March were comparable at all stations, where PM_{10} at Jongno, Jung, and Jongno Road ranged between 45-47µg/m^3 in February, and 45-49µg/m^3 in March, while Seoul station had 63.1µg/m^3 in February and 64.6µg/m^3 in March 2018. Jongno Road reported amounts just 1-2µg/m^3 higher than the background stations which was unexpectedly low for a roadside station during the same period. It may be that the installed location was not reflecting the PM_{10} level which it should have done or the extraction method (counting the concentration of particles) is not reporting adequately (Lee and Park, 2019). However, if nothing has gone wrong with the measurements, the result from roadside stations indicated a wide variation based on the installation or perhaps traffic volumes, which even might be comparable to the background level.

[6]walking, passenger car, carpooled people, city bus, regional bus, town bus, intercity bus, express bus, other buses, subway, rail, speed rail, taxi, small freight, large freight, motorbike, bicycle, and others

Table 5.1 Summary statistics of PM_{10} concentration at four stations in January-March 2018 (Units: µg/m^3)

Type	Month	Background		Roadside	
Stn		Jongno	Jung	JongnoRd	Seoul Stn
Mean		44.8	44.6	47.3	58.2
sd	January	24.3	24.9	26	33.7
Min		11	12	10	6
Med		38	38	41	50
Max		171	180	183	247
Mean		46.1	45.3	47.1	63.1
sd		22	21.9	22.7	29.7
Min	February	5	5	5	7
Med		40	38	40	56
Max		144	142	146	176
Mean		46.3	45.1	49.9	64.6
sd		30.7	29.9	31.5	35.7
Min	March	3	4	5	5
Med		42	41	44	61
Max		142	146	150	176

Despite a monthly average maintaining between 40-60µg/m^3, major PM_{10} episodes were observed across all sites in mid-January, early and late February, and in mid and late March (see Table 5.1). PM_{10} levels tend to fluctuate day by day rather than from day-to-night hours (see Figure 5.5). This may be due to the low wind speed preventing particles from being transported outside the urban area, with vehicles contributing extra particles from tyre and brake wear to the local atmosphere (Panko et al., 2013).

Furthermore, South Korea's PM_{10} guideline of 100µg/m^3 was exceeded by 302 hours at Seoul Station (14%), 133 hours at Jongno roadside (6.1%), 117 hours at Jongno background (5.3%), and 104 hours at Jung background (4.8%) (see Table 5.2). Although the hours vary by stations, March had the most counts of extreme PM_{10} that exceeded 100µg/m^3.

Traffic in Seoul CBD

Figure 5.6 illustrates the average vehicle counts by the hour of the day between January-March 2018. The figures are the summation of traffic in both directions. As seen from

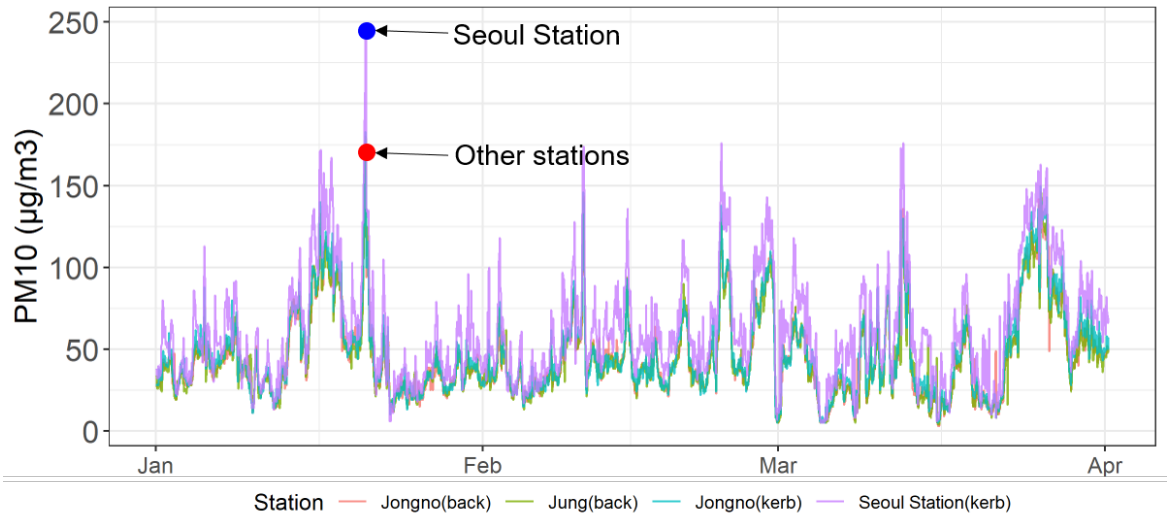


Fig. 5.5 A comparison of Hourly measured PM_{10} between the background and roadside stations. The smooth lines (rolling average) inform that the Seoul station (purple) was 16µg/m^3 higher than the other stations. The largest difference was observed on January the 20th, with 247µg/m^3 reported on the kerbside and 171µg/m^3 on the background.

Table 5.2 Counts of a number of times that exceeded the 100µg/m^3 threshold in every hour during January-March 2018 (Units: µg/m^3)

Month	Jongno(back)	Jung(back)	Jongno(road)	Seoul Station(kerb)
Jan	34	34	42	88
Feb	21	18	25	91
Mar	62	52	66	123
Total	117	104	133	302

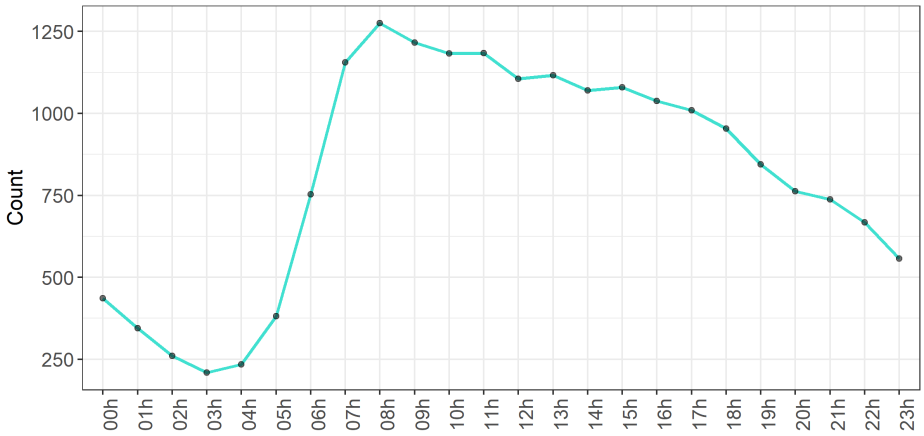


Fig. 5.6 Average traffic count of the CBD stations by the hours of the day (January-March, 2018). The numbers were added from both directions.

the hourly data during the study period, a rapid increase occurred between 4:00-8:00 where the numbers soared from 250 to 1250 each hour. The count slowly decreased from 9:00 but maintained above 1000 cars per hour until 17:00; then continued to decline until 3:00.

To understand the volume of the traffic by roads, the total counts were decomposed by each road (see Figure 5.7). In Figure 5.7, Sajik and Dongho roads have the highest volume of all with over 2000 during the working hours, followed by Jongno, Samunan and Seosomun roads that had over 1100 vehicles per hour at the same time. Sopa and Daesagwan were the quieter roads in the CBD that mostly remained at around 200-500 vehicles per hour.

OD Matrix

Prior to assigning fractional OD matrices to pedestrian and driving agents, this section explored the geodemographics of trip origins and destinations based on the mode of transport. In Table 5.3, the daily inbound passengers in 2015 comprised subways (749,165 persons, 34.4%), buses (all types, 469,691 persons, 21.6%), personal vehicles (367,723 persons, 16.9%), taxis (235,903 persons, 10.8%) and so on. Passengers also selected the subway (730,361 persons, 34.37%), buses (413,949, 19.48%), passenger cars (380,374, 17.9%), and taxis (218,874, 10.3%) for their outbound journey. Most of those who travelled within the CBD tend to walk (135,508 persons, 32.6%), take buses

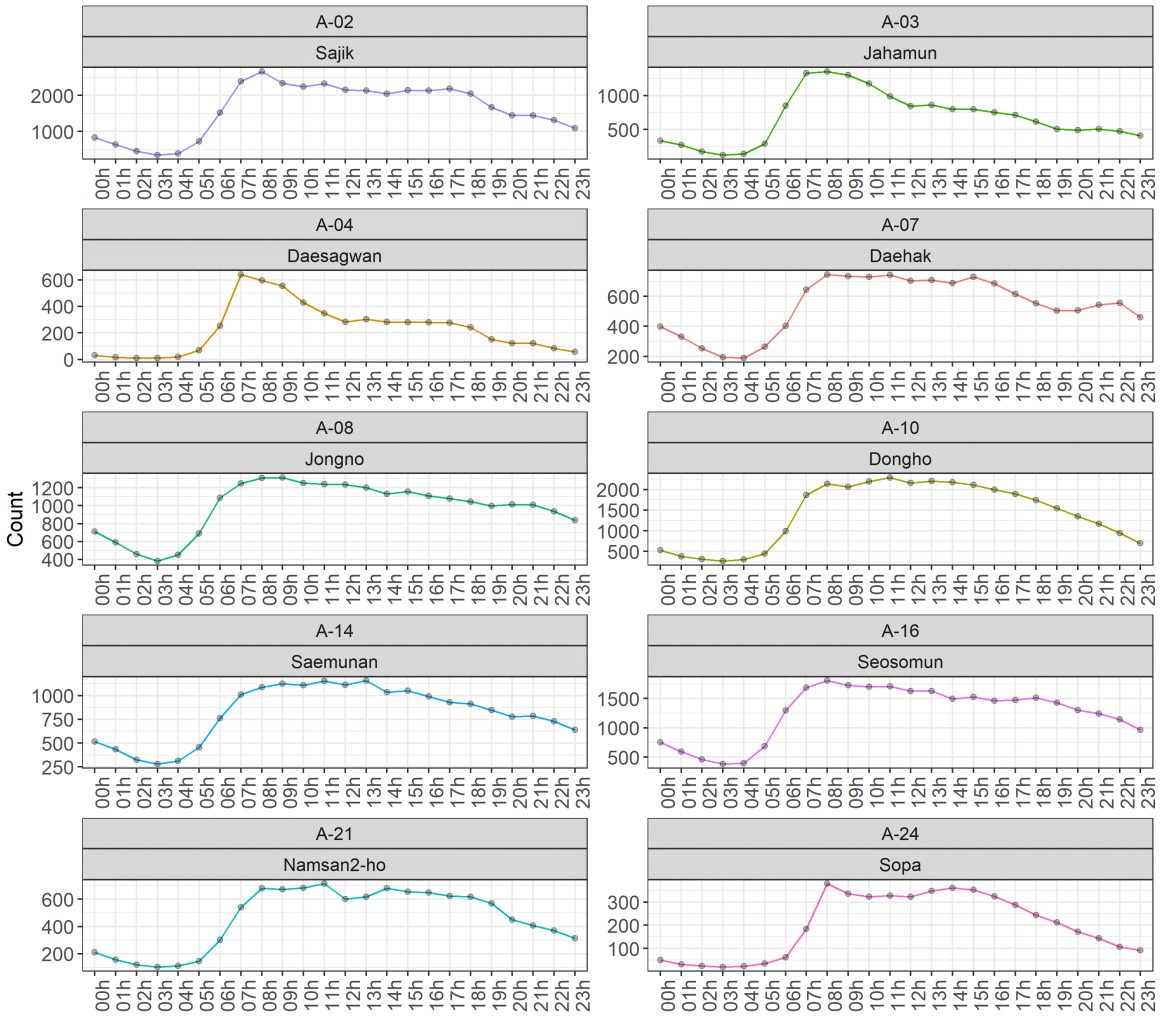


Fig. 5.7 Hourly traffic counts by 10 CBD road checkpoints

(105,676 persons, 25.4%), or subways (83,221 persons, 20%). In summary, 34% of the passengers used subways to get to and leave from the CBD, while 32.6% of people who travelled within the CBD walked to their destination.

It is informative, before conducting an ABM simulation, to understand the fraction of the population moving to certain areas, although it is unknown where the people's workplaces are. The fraction of OD data simply filters the counts of passenger cars and converts to a scale of 0-1. As mentioned earlier, there is a mismatch of administrative codes between the year the vehicle counts were collected and the year that the OD matrix was generated, in other words, some admin codes were in 2015 are missing

Table 5.3 Rank of transport modes for daily CBD travel: inbound (heading to CBD), outbound (heading back to origin), and travelling within CBD

Rank	Direction	Type	Persons (%)
1	Inbound	Subway	749,165 (34.4%)
2		Buses	469,691 (21.6%)
3		Cars	367,723 (16.9%)
4		Taxis	235,903 (10.8%)
1	Outbound	Subway	730,361 (34.4%)
2		Buses	413,949 (19.5%)
3		Cars	380,374 (17.9%)
4		Taxis	218,874 (10.3%)
1	Within CBD	Walking	135,508 (32.6%)
2		Buses	105,676 (25.4%)
3		Subway	83,221 (20%)

in 2018. The result is shown in Table 5.4. This matrix is the final version that will allocate the sample of vehicles to their destinations.

It is worth raising the caveat of using a fractional OD matrix that matches the data too exactly when using a subsample of the population: an accumulation error may occur that gives a sum more than the total number of agents that exist. In mathematics, the so-called rounding error or floating-point precision error occurs when *"the result produced by a given algorithm using exact arithmetic and the result produced by the same algorithm using finite-precision, rounded arithmetic."* (Ueberhuber, 2012). In this model, this occurs when the counts of the allocated agents (by origin/destination) are larger than the counts of the agents. The opposite would not be a problem because the agents can be given another arbitrary destination or perhaps be removed.

For example, 33 agents of origin 1101053 will be allocated to each sub-district including itself. The model will calculate 33×0.19, then 33×0.01 (see Table 5.4), and so on, which will total 30. As mentioned, the three agents that do not have an allocation can be allocated to an arbitrary destination but were removed for this study. However, in two sub-districts, 1101060 and 1102058, because each sub-district had 10 agents, but the arithmetic results returned 11 agents and 13 agents. To avoid these computational problems float values were reduced until the error disappeared. The adjustment might seem arbitrary but is necessary for model consistency and the difference of agents would not be too problematic as it would not noticeably change the emission levels.

Table 5.4 A fractional OD matrix for resident drivers. The first column is the admin codes of origins and the first row is the admin codes used to allocate destinations

Code	1101053	1101054	1101060	1101061	1101063	1101064	1101072	1101073	1102052	1102054	1102055	1102057	1102058	1102059	1102060
1101053	0.19	0	0.01	0.2	0.01	0	0.37	0.01	0.01	0.07	0.09	0	0	0	0.04
1101054	0	0.29	0.06	0.14	0	0.06	0.3	0.07	0.03	0.04	0	0	0	0	0
1101060	0	0.07	0.17	0.3	0.07	0	0.07	0	0.05	0	0	0.04	0	0.06	0
1101061	0.06	0.04	0.07	0.48	0.02	0	0.02	0.02	0.02	0.03	0.05	0.01	0.02	0.12	0.02
1101063	0.05	0	0	0.15	0.14	0.04	0.2	0	0.05	0.1	0.22	0	0	0.03	0.02
1101072	0.34	0.01	0.02	0.1	0.01	0.06	0.36	0.05	0.01	0.01	0.01	0	0	0	0.01
1101073	0	0.08	0	0.03	0	0.06	0.01	0.5	0.15	0.08	0.01	0	0	0.08	0
1102052	0.04	0	0.01	0.11	0	0.01	0.01	0.06	0.3	0.1	0.26	0.02	0.05	0.02	0.01
1102054	0.13	0.01	0	0.07	0.01	0	0.01	0.08	0.03	0.23	0.27	0.01	0.02	0.01	0.11
1102055	0.13	0.01	0	0.05	0.03	0.01	0.09	0.03	0.19	0.05	0.26	0.02	0.01	0.07	0.06
1102057	0	0	0.01	0.04	0	0.03	0	0	0.11	0.34	0.23	0.13	0.01	0.07	0.03
1102058	0	0	0	0.2	0	0	0	0	0.16	0.17	0.07	0.05	0.16	0.04	0.03
1102059	0	0.01	0.01	0.15	0.04	0	0	0	0.02	0.19	0.25	0.06	0.01	0.23	0.02
1102060	0.04	0	0.01	0.1	0.02	0	0	0.01	0	0.17	0.05	0.06	0.07	0.05	0.42

5.4 Agent-based modelling: Setup

This section documents the essential settings of the agent-based model based on the new ODD protocol (Grimm et al., 2020). The difference between the first version (Grimm et al., 2010) is the deprecation of the "Detail" section, and the suggestion to attach codes to the document to avoid any ambiguity. The remainder of this section follows the new ODD protocol structure: Section 5.4.1 Model Purpose, Section 5.4.3 Entities, State Variables, and Scales, Section 5.4.4 Process Overview and Scheduling, Section 5.4.5 Design Concepts, Section 5.4.2 Initialisation, Section 5.4.6 Input Data, Section 5.4.7 Sub-Models.

5.4.1 Model Purpose

The purpose of this model is to understand commuter's exposure to non-exhaust PM_{10} emissions, and to make a preliminary estimate of their health effects. Specifically, this study addresses the following questions: (1) how does health risk vary between resident drivers and subway commuters? (2) how might vehicle restriction plans for cars mitigate the level of exposure and air quality in the CBD? and (3) how might awareness of pollution levels (e.g. via phone applications) reduce the risk of exposure?

The model illustrates the following patterns: (1) the fraction of population at risk by mode of transport and (2) the total numbers of traffic and pollution levels by road in a context that is representative of realistic conditions in the Seoul CDB.

Pattern 1: Population at risk by the mode of transport

- This pattern reflects how an individual's health might deteriorate from PM_{10} exposure depending on the mode of transport they take, and how much time is spent under extreme PM_{10} conditions. Health decline occurs when PM_{10} exceeds the 100µg/m^3 level: a nominal health index is used, starting at 300, and individuals are labelled as "at risk" if the value drops below 100[7]. The population at risk is a fraction of individuals with a health value less than 100 relative to the total population.

Pattern 2: Traffic load and pollution concentration

[7]The same setting was already implemented in the previous chapter

- This pattern emphasises the spatial variation of the pollution attempts to understand how one road is polluted relative to other roads, and how much traffic contributed to that. In other words, the commuting patterns and traffic flow generate some fraction of the emissions that impact people's health. This potentially allows a feedback between pollution and behaviour to be simulated. Although only a fraction of vehicles is represented, we can use sensitivity studies to test how important this might be to the realism of the output.

5.4.2 Initialisation

Vehicles

A) *Resident vehicles*: 399 resident vehicles were sampled and imported in the model. The vehicles accounted for 10% of the total vehicles registered in each sub-district. Having tested different sample sizes ranging from 10-70%, the pollution level or health effect of 70% samples was not significantly different to the 10% sample size.

During weekdays, trips are made along the shortest path and will not change throughout the simulation, while the weekend journeys are random. During each trip, the vehicles will keep some distance from other vehicles. Vehicles contribute pollution to the surrounding atmosphere as they move along the streets. Details about trips, emissions, and exposure are explained in the following sections.

B) *Non-Resident Cars*: Unlike resident cars, non-resident cars do not have any specific navigation aims, but rather, play a role as pollution-generators inside the study domain. The vehicles will follow traffic signals and keep their distance from the vehicles in front, but however, will be removed completely when they reach the end of the road (at the domain boundary, see Figure 5.12). The randomness of travel directions is to simulate general movement during the vehicle's time in the CBD, in the absence of more detailed data. These vehicles are not present during the model settings but will appear when the model is executed.

People

C) *Resident drivers (tied with resident cars)*: The drivers are tied with cars but do not move nor appear on the interface. This is to improve the model running speed and

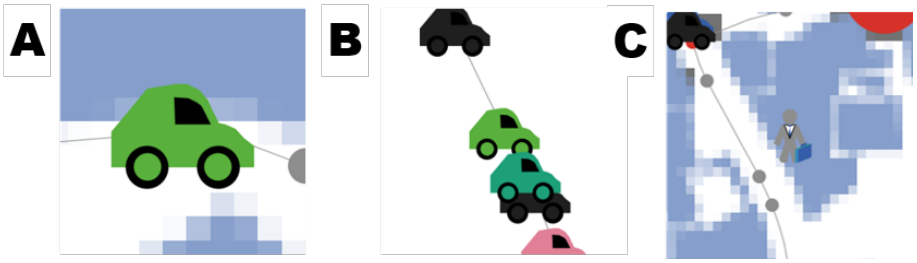


Fig. 5.8 Agent types used in the simulation. Type A is a resident vehicle (tied with a driver), Type B is a non-resident vehicle, and Type C is a pedestrian commuting by metro

to prevent any computational errors between linking and unlinking cars from people. The drivers lose health when they are instantaneously exposed to the nominal PM_{10} threshold of 100µg/m^3.

D) *Subway commuters*: To execute the model efficiently, the model populated 1932 persons (1% of the subway commuters) and gave each agent a random building location within 10 patches from the subway entrance. Once the location was assigned, the agents are asked to walk to their offices based on the local search algorithm (see details in Section 5.4.7).

5.4.3 Entities, state variables, and scales

As mentioned in the previous section, this model consists of three types of mobile agents (resident cars with drivers, non-resident cars, and subway commuters) and two types of fixed agents (traffic signals and entry points where the vehicles are fed into the study area). Since all mobile agents were introduced in the previous section, this section adds state variables as well as the introduction of fixed agents.

The state variables for the mobile agents, vehicles and people, are documented in Table 5.5 and Table 5.6. Resident vehicles have their origin and destination both in patches and nodes, where `home patches` and `destination patches` are considered as indoor spaces that require PM_{10} to be adjusted to the indoor level (Kreider et al., 2019). The indoor/outdoor ratio is further explained in Section 5.4.5. `path-work` and `path-home` provide links between home and the destination node, and the `positive direction` guides vehicles to follow the links of `path-work`, whereas the `negative direction` guides vehicles to follow that of `path-home`. `Link-counter` answers the question, "How

Table 5.5 Vehicles in the model have state variables related to their trip

Code	**Description**	**Example**
fueltype	Type of fuel	"Gasoline"
origin	One of nodes set as origin	(node 1)
destination	Patch set as destination	(patch 40 40)
goal	The closest node from destined patch	(node 2)
path-work	List of nodes between home and work	[(node 1) (node 2)]
path-home	List of nodes between work and home	[(node 2) (node 1)]
nodes-remaining	Number of nodes from the list	24
myroad	List of roads between home and work	[(link 1 2) (link 2 3)]
current-link	Current road	(link 1 2)
district_name	Name of district	"Myungdong"
district_code	Code of district	1102055
link-counter	Cumulative counter to arrival	0
direction	Direction to work (1); to home(-1)	1
time-at-work	Minutes at work	524
random-car	Boolean of random / resident cars	True/false
parked	If the vehicle is parked	True/false

many links before the vehicle stops?" Incrementing by 1, the `link-counter` will stop when it meets the `nodes-remaining` value, and then `parked` changes from false to true. After spending time-at-work for 480+α minutes ($\alpha < 60$), the cars will start the journey back home.

For subway commuters, as they come out of the subway entrance, known as the origin, they walk to their goal patch using the shortest distance, when the awareness scenarios are not activated. If awareness is activated, the individual moves to one of the three patches on the direction that has the lowest PM_{10}. `Heuristics` is the distance between the origin and goal, which will decrease as the individual moves towards the `goal` patch. As soon as the individual reaches the goal, the `arrival?` status will change to TRUE. Note that for some agents whose `Heuristics` is less than 1 and less than the walking speed will be stuck at that location. To avoid the error, the individuals whose `Heuristics` is less than 1 will automatically move to the office location and will convert their `arrival?` to TRUE and start working. As with resident vehicles, `time-at-work` shows the remainder of the working time. For example, if an individual arrived with 500 minutes the minutes will decrease and the departure will be made once the minutes reach zero. For visual purposes, this model temporarily

Table 5.6 People in the model have state variables related to their trip

Code	Description	Example
origin	Subway node set as origin	(exit 20)
origin_patch	Patch of origin	(patch 20 20)
goal	Patch set as destination	(patch 40 40)
current	Current patch	(patch 30 30)
Heuristics	Distance between current and goal	0.11
arrived?	Whether they arrived to their workplace or not	T / F
time-at-work	Number of ticks spent after arrived?	480
direction	Direction to work (1), stay (0), and home (-1)	1/0/-1
arrive-tick	Ticks spent between exit and arrived?	39
Health	Nominal health level (starts from 300)	275
Hour	Hour when the agent arrives at the subway station	7
Minute	Minute when the agent arrives at the subway station	48

Table 5.7 Variables of a traffic signal

Code	Description	Example
Dong_code	Code of admin (provided by Census)	1102055
Intersection	Boolean of road intersections	T/F
Auto?	A timer to change signals	15
Green-light?	Boolean of green lights	T/F

removes the workers whose arrive-tick is over 80 so that the display is less cluttered, but makes them re-appear after work.

Traffic signals are arbitrarily created at junctions that consist of three roads or more (see Table 5.7). More traffic lights are installed between road segments in the real world, but the intention here is to control the traffic and to articulate the resuspension of dust – the emitted PM_{10} will remain a little longer due to the traffic queues in front of the traffic signal. `Intersection` is a Boolean variable that determines whether they have three roads to become qualified. The duration of red and green signals are determined by a timer variable termed Auto?. If `Auto?` is over five (ticks) and the traffic signal is coloured green, the `Green-light?` will become true and vice versa.

Subway entrances are set as an origin for subway commuters (see Table 5.8). There are 26 subway stations in the study area with line numbers 1, 2, 3, 4, and 5, and coded as `s_entrance`.

Table 5.8 Variables of a traffic signal

Code	Description	Example
Line	Line number of Seoul Metro	1

5.4.4 Process overview and scheduling

The model runs at ***a one-minute time step***, and variables are collectively updated until the simulation terminates (see Figure 5.9). The simulation starts at 7:00am on January 1st, and ends at 23:59 on March 31st 2018. The diagram shows the journey of vehicles and humans, and where the vehicles produce pollution (see Figure 5.9 process a and Figure 5.10), the agents who are exposed to over 100µg/m^3 of ambient PM_{10} in the study area are expected to have their health decreased (see Figure 5.9 process b and Figure 5.11). Although the full journey to the CBD is not simulated in this study, subway commuters are assumed to be exposed to the ambient level PM_{10} between early morning and late in the evening even if they do not appear on the interface. The cumulative updates of the risk population and the PM_{10} concentration by roads are exported to a single spreadsheet at the end of the simulation.

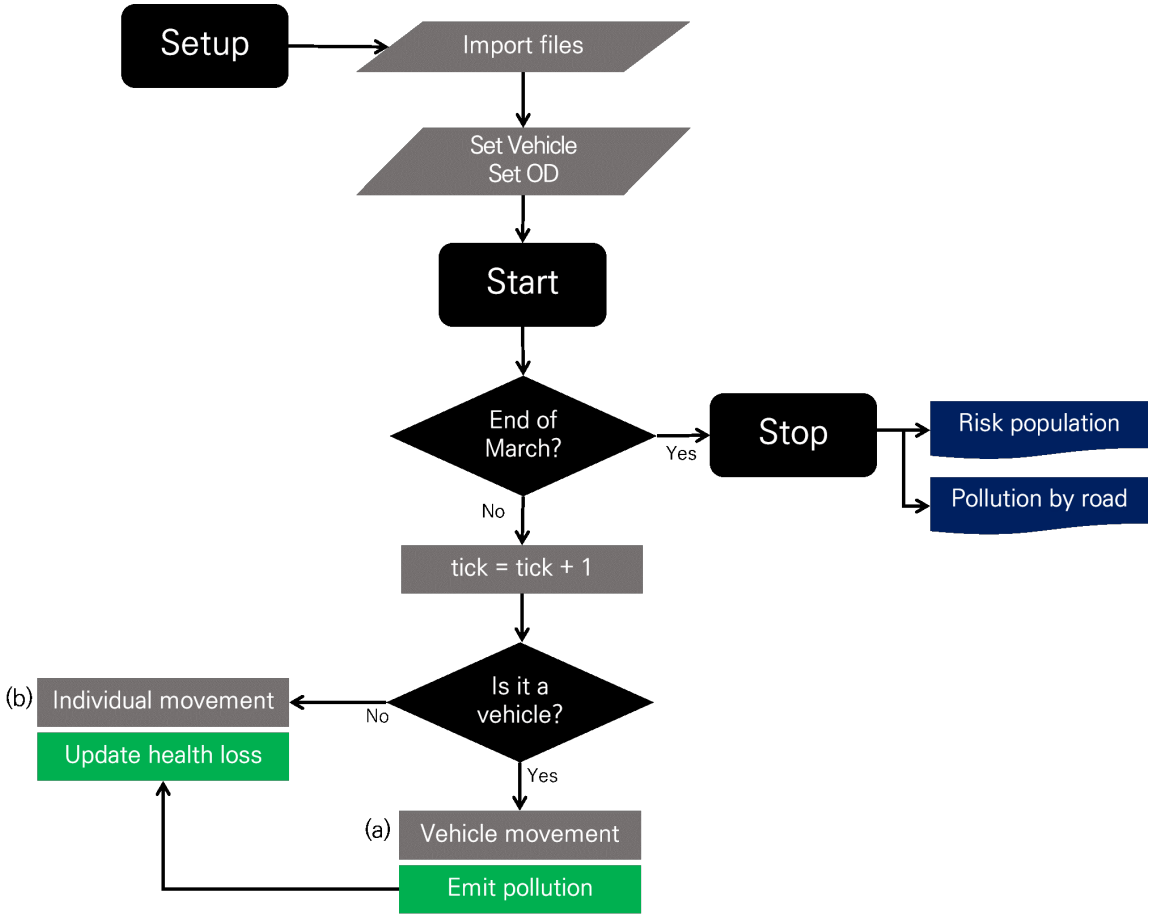


Fig. 5.9 A nested flow diagram describing the behaviour of agents and their landscape. During the setup period, buildings, pollution and roads are created. Registered vehicles are also created with their allocated destinations. The model starts at 6:00am on January 1st, and ends at 23:59 on March 31st 2018. Each tick will count as one minute. If the agent is a vehicle, then it follows the behaviour of a vehicle (see Figure 5.11 for details); otherwise, it follows the behaviour of an individual (see Figure 5.10 for details). If the simulation stops, then it will print the population at risk and pollution levels by road.

Vehicles

Vehicles are divided into two groups: 1) resident vehicles or 2) vehicles with random movement (see Figure 5.10 for details). The driver's health loss will be explained in the later section.

Vehicles in general:

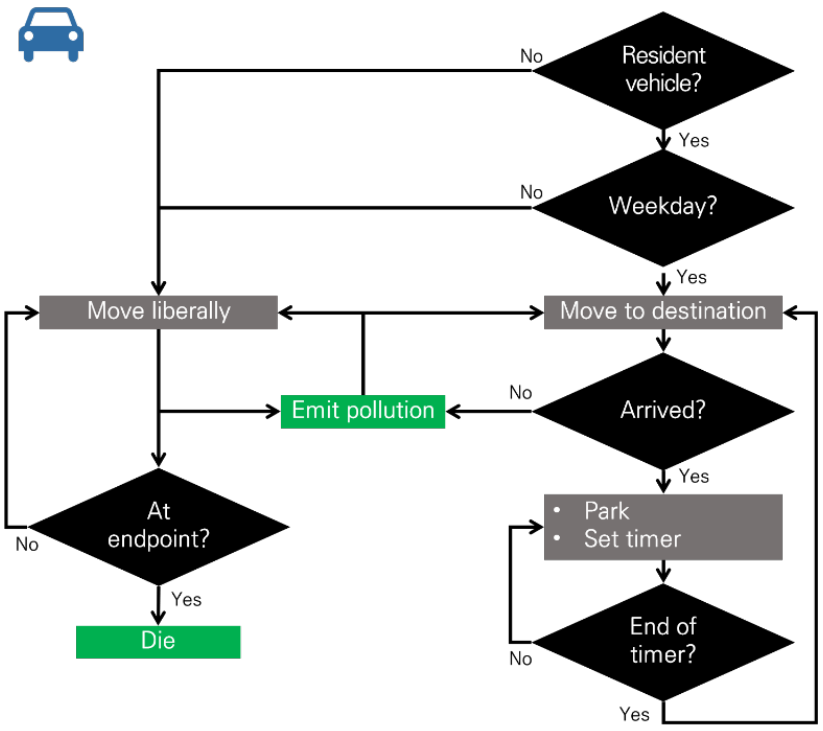


Fig. 5.10 Flow chart for resident and incoming vehicles. If the vehicle's owner is a CBD resident, the vehicles will move to their assigned destination. The vehicle will emit pollution until it ends the journey. As the vehicle parks at the destination, the timer will start to countdown from $480+\alpha(\alpha<60$ mins) to 0 mins and will head back home once the timer reaches zero. If the vehicle is non-resident, it will move generally and disperse non-exhaust pollution until it leaves the domain.

- Both vehicle profiles maintain a safety distance of 1 patch ($\approx$ 30m) between themselves and the vehicle in front. During the journey, vehicles will pollute and disperse non-exhaust PM_{10}, regardless of fuel types. Vehicles are asked to stop in front of the "Red" traffic signal. More information regarding the pathfinding algorithm and PM_{10} pollution will be introduced in the Sub-Model section.

Resident Vehicles:

- Vehicles will move across road networks to their destination node, stop during office hours, and head back to the `origin (node)` again using the same route during weekdays, but will move away from the study area over the weekends for non-working activities, e.g. shopping, weekend journeys, and places to worship. During weekdays, each vehicle will stop the journey if the vehicle has arrived

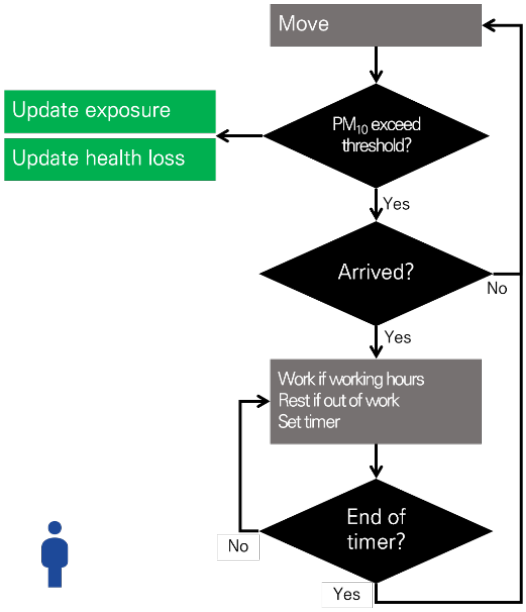


Fig. 5.11 Flow chart of a subway commuter's journey. While the person is walking, it's health will degrade when the PM_{10} is above 100. If arrived, the person will stay until the timer ends and head back to the station..

at its destination node. After its arrival, the state variable, timer, counts down from $\geq$480mins, and as soon as the timer reaches zero, the vehicle will head back home. Extra time from 0 to 59 minutes is given to all agents assuming agents walking to car parks or spending additional time to wrap up their work. Each vehicle has a driver whose health will decline if the PM_{10} inside the vehicle is over 100µg/m^3.

Vehicles with random movement:

- Vehicles are assumed to have come from the outside. These incoming vehicles make trips to any areas inside the CBD, generating vehicles from the hourly traffic data. Since the spatial extent is restricted to the CBD zone, this model made the outbound cars disappear at any endpoints of the road network. Since the model had a limited capacity of vehicles (~2,500), the traffic count was further decomposed by 5% on the scenario, as well as 2.5%, 10%, and 20% on the sensitivity experiment. Note that if a vehicle checkpoint station had less than 1,200 vehicles in an hour, then a 5% sample would not feed in any vehicles for

that hour, but this was not a problem since not a large difference was seen in between the ratios – details are demonstrated in the Sensitivity section.

The basic code for the vehicle's movement was based on the Venice model (unpublished, but the source code was shared in 2017: https://www.gisagents.org/2009/02/agent-based-models-for-venice.html).

Subway Commuters

When the simulation commences, the subway commuters are transported to the subway entrances at the hour and minute they have on their state variables. Once the agents arrive at their subway entrances, they walk to their destination buildings using the shortest distance regardless of the pollution levels. However, if the awareness scenario is activated, they will navigate following the lowest PM_{10} to their destinations.

While walking, everyone gets an equal chance of exposure to the PM_{10} threshold, but the degree of health loss will depend on how much time is spent outside when the PM_{10} is over 100, and how long the distance is between the subway entrance and the agents office. Moreover, the agents whose office is adjacent to roads might lose more health because the pollution generated from the roadside can affect the indoor pollution, e.g. opening and closing windows (Kreider et al., 2019).

Traffic Signals

When the simulation starts, each signal will be given a random number between 0 and 11 and will count down to 0. Between 5-10 is the red light that allows the vehicles to pass, and 0-4 stops the vehicles. The timer will reset to a random number again once the counter reaches 0. This will give full randomness to the traffic signals in the study area.

Subway Entrances

As the simulation commences, the model chooses 4 out of 26 random stations to create commuters. It will be a returning point for commuters to travel home.

5.4.5 Design Concepts

Basic Principles

This exposure model was developed to illustrate how the population in the CBD zone can be exposed and possibly lose health in response to non-exhaust PM_{10} emissions. There is extensive literature on traffic-related exposure, mainly associated with NO_x emissions, or with population exposure to NO_x (Beevers et al., 2013; Halonen et al., 2016; Min et al., 2020), but not with non-exhaust emissions of particles. With increasing awareness that non-exhaust emissions are important (Air Quality Expert Group, 2019; Amato, 2018), this study builds a health impact assessment model based on non-exhaust PM_{10} emissions.

The rationale is that the particles generated by non-exhaust emissions (i.e. tyre and road wear particles) have been problematic for many years (Kreider et al., 2019), but despite new vehicle models that comply with the environmental regulation, the percentage of non-exhaust emissions are increasing in many countries (Air Quality Expert Group, 2019; Al-Thani et al., 2020), and population health may be under a serious threat from instantaneous pollution rise. As experts raise concerns about the potential threat that the non-exhaust particles can bring to the local atmosphere, there should be a preparation for further regulations to non-exhaust particles in the near future (Air Quality Expert Group, 2019).

As a starting point, the model asked resident and non-resident vehicles to generate and disperse PM_{10} to the local atmosphere, namely on road and nearby pavements, while subway commuters and drivers are the susceptible individuals who are exposed to PM_{10} emissions. On the other hand, the background PM_{10} generated the value from the urban monitoring stations within the study domain. Each agent group has different behavioural patterns, which was explained in the previous section.

Emergence

The percentage of the population at risk (i.e. those with health under 100) emerges from a balance between exposure to a PM_{10} threshold of $100\mu g/m^3$ and recovery. In practice, the emergence can be an acute response to PM_{10} exposure before the natural recovery begins to take effect. The emergence pattern will differ by which means of transport the individual is commuting with. This is because subway commuters

are exposed to the ambient atmosphere during their walk from subway entrances to offices, while resident drivers spend most of their time indoors or in transit but have a higher chance of inhaling polluted air from road traffic. Although findings that extreme particulates were even higher than other transport modes have been reported (Mao et al., 2019; Smith et al., 2020), this study omitted the journey of subway commuters because information of the start and end stations are not provided in the OD data, which is very crucial for microscopic modelling.

Adaptation

This study has two aspects of adaptation: pathfinding and health recovery. With regard to pathfinding, the subway commuters either walk along the shortest path when the awareness scenario is deactivated or find the best way to avoid high-polluted locations of PM_{10} exceeding over 100µg/m^3 when the awareness scenario is activated. If the awareness scenario is activated but the agent struggles to find a path below 100µg/m^3, the agent will then find the lowest PM_{10} of the possible routes and move to that location. Resident drivers have their health deteriorate when the patch on the road is at least 144µg/m^3 because the indoor-outdoor ratio between inside-vehicle and ambient air is 0.7. If the awareness scenario is activated, the driver will take a free trip during weekends – at the beginning of Saturday or Sunday – and only when the driver's health is over 100. Conversely, if the awareness scenario is not activated, the drivers will take a trip regardless of their health. Both groups have their health recovered by the same amount at a nominal value of 10 out of 300 per timestep.

Sensing

Subway commuters are exposed to the PM_{10} at which they are located. If the PM_{10} is over 100, the commuters will lose health according to the health loss equation. Subway commuters also use the shortest distance to their workplace when the awareness scenario is not activating or find the lowest value of PM_{10} amongst the front three patches in the direction they are moving. Additionally, everyone has its own time of arrival at the subway station. For instance, if the hour and minute variable of agent X is 8 and 12, agent X will appear at the station at 8:12 am. Both subway commuters and drivers have fixed working hours with a few minutes of extra time (up to one hour) to finish the daily work. The extra minutes differ every day.

The vehicles can sense one radius distance between the vehicles in front and behind and the traffic signals. As with subway commuters, drivers also have their destination time to work. After departure, the vehicles travel on the shortest route to their workplace.

Interaction

Interactions occur between the PM_{10} levels and the agent's health. That is, subway commuters who are exposed to over 100µg/m^3 of ambient PM_{10} on the current patch will lose health, while the drivers will lose health according to the non-exhaust emissions from vehicles. Vehicles interact with traffic signals. The vehicles stop in front of the red lights and start when the light changes to green.

Stochasticity

Vehicles

- Vehicles have different origin and destination locations at every setup.
- Resident vehicles park for 480 mins (ticks) with a random number of extra numbers (up to 60).
- A vehicle has a minimum speed of 0.5 patches per tick and a maximum speed of 3.5. In cases of queueing, the deceleration ranges between 0-0.7 and the acceleration ranges between 0-0.5.
- Non-resident vehicles are fed into the study area according to the traffic monitoring statistics; however, the direction and time spent are random. Since the model has a limited capacity of vehicle numbers, a randomly selected 0.1% of the vehicles will disappear every minute between 10pm and 4am and 0.25% during the rest of the hours. This is to assume that the vehicles have driven out of the CBD[8]. For example, if there were 2000 vehicles in the study area at 10am, five vehicles[9] will disappear, and four vehicles in the next minute.
- Resident vehicles will select a random road to travel outside of the CBD.

[8]Having tried multiple ways to induce the non-resident vehicles outside, the most effective method was to eliminate a random set of vehicles

[9]$2000 \times .0025 = 5$

Subway Commuters

- In the setup process, subway commuters choose a random subway station, then assign one of the buildings within 10 radii as their workplace.

Health loss and recovery

- When a human agent is exposed to PM_{10} over 100µg/m^3, the health loss equation subtracts the amount of health based on the factor α, where α ranges between 0 and 0.2. The parameters are tested for sensitivity, but only one parameter is used for scenario forecasting. See Section "Health Loss and Recovery" for details (p.232).
- Infiltration ratio (indoor/outdoor ratio) varies by the microenvironment and the time spent. This study estimates the infiltration from the ambient PM_{10} of the current patch to indoor spaces such as houses, workplaces, and transits (Kreider et al., 2019). Ratios for each microenvironment compared to the outdoors are described as follows:
  - Houses: 0.2-0.7 (Kreider et al., 2019; Leung, 2015). The home patch discounts the ambient PM_{10} by an index between 0.2 and 0.7.
  - Workplaces: 0.2 (Kreider et al., 2019). The work patch discounts the ambient PM_{10} by 0.2.
  - Vehicle: 0.7 (Kreider et al., 2019). The patch where the vehicle is stopped decreases by 0.7 of the ambient PM_{10}.
- Health recovery is stochastic at the assumption that one can recover better than another. Any agent whose health is below 100 and remains at a stable place (home/office) will recover by $10 + \epsilon$ per minute (ϵ being between 10 and 20), until its health returns to the 'non-risk' state.

Observation

As the first research question asks whether the risk rate varies by groups, the graphical output of the model shows the risk rate of subway commuters and resident drivers by time. Date, hour, and time are displayed on the interface to inform the current time. The average PM_{10} and a few road points are collectively monitored until the simulation

ends. Subway commuters will not appear on weekends but will still be exposed to PM_{10}. Resident drivers, on the other hand, will travel in a random direction during weekends, but once they reach the end of the road they will stop until 10pm and return to their origin. The returning procedure moves cars directly to their origin, which is intended to simplify the process.

Subway commuters and resident vehicles do not travel to work on weekends and national holidays. So, the interface will look less busy on Saturdays, Sundays, Lunar New Year (February 15-17th), and Independence Movement Day (March 1st).

5.4.6 Input Data

Area: The CBD area (16.7km^2) comprises two districts of Seoul, namely Jongno and Jung. Jongno has 8 sub-districts and Jung has 7 sub-districts. This model used 30m resolution, making 155 horizontal and 192 vertical patches.

Roads: Road layer is the most important component to simulate vehicle trips along with the road network and emit pollution. Roads in the real-world have different hierarchies (e.g. two lanes, four lanes, eight lanes), but this study started with a single type of road.

Buildings: Buildings are used as the agent's office places which are brought in from OpenStreetMap. When the model is loaded, commuters are allocated their own building ID through a random allotment.

Hourly Traffic: Hourly traffic was provided from the Traffic Monitoring Department affiliated with Seoul Metropolitan Government (see entrances from Figure 5.4). Figure 5.6 presents the traffic counts from the simulation by the monthly and weekly period. It is seen that the traffic volume declines on weekends, and on the Lunar New Year holiday (February 15-17th).

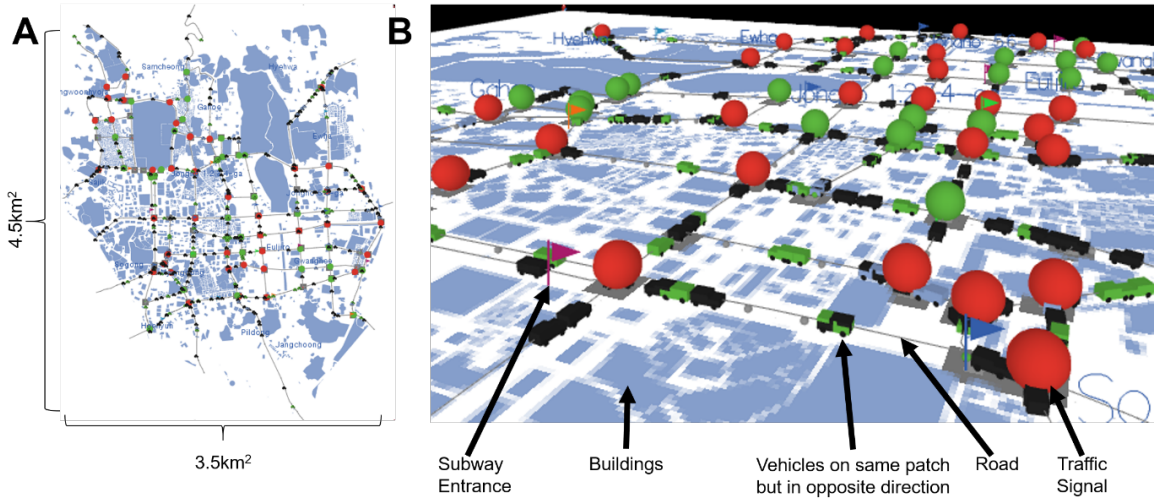


Fig. 5.12 Study area in 2D (A) and 3D (B)

5.4.7 Sub-models

The Pathfinding Algorithm: A* and the Local Search Algorithm

In line with the previous chapter, this study also assigns the agent's origins and destinations according to the Origin-Destination Matrix; however, the way of application is slightly different. As a quick repetition, the previous chapter used the matrix to choose a fraction of the population from their origins and allocate it to their destinations. The fraction of the population that was allocated outside the study area was not considered for further measurement. This chapter considers the same mechanism by gathering the population of the origin and assigning destinations based on the fraction assigned in the matrix. While the previous chapter only considered agents moving between their origins and destinations, this chapter specifies the resident vehicles to follow the shortest route between their origins and destinations on the road link.

Once an agent's origin and destination are assigned, the mechanism to connect the two points requires a pathfinding algorithm. This study gives a separate algorithm to vehicles and people. For vehicles, the model used A*[10] (Zeng and Church, 2009). A* calculates the lowest cost distance from its origin and destination and traces the path where the cost is smaller. This can be formulated as:

[10] A* is one of the most popular path-finding algorithms together with Dijkstra's from their vertices and segments, which in real life may represent road networks.

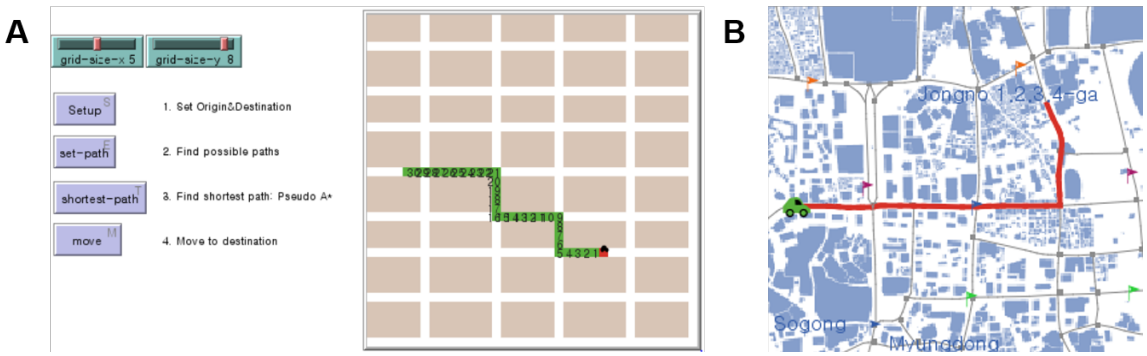


Fig. 5.13 (A) is a sample of an agent finding the shortest path from the origin (red patch) to its destination (light green patch), and(B) is the application of the shortest distance on link data

$$f(s) = g(s) + h(s), \tag{5.1}$$

where `s` is the state, `g(s)` is the cost from the origin to the current `s`, and `h(s)` is the heuristic estimation between the current state and destination, which adds up to the total cost at `f(s)`. In this context, an individual's heuristic measurement is referred to as the shortest Euclidean distance to its destination. The A* algorithm is based upon Dijkstra's algorithm but uses the heuristic framework to shorten the calculation time and optimise the shortest path.

The interface below is an example of an A* algorithm (see Figure 5.13A, B). Figure 5.13A indicates a gridded guideway between the vehicle's origin (red) and destination (green). set-path finds all the steps from the origin patch to all possible steps inside the virtual world. The code will colour the road green and add a step label on the path. Among the steps, the shortest-path section traces the shortest grids as per the A* method, which then allows the vehicle along the path with the move section. Figure 5.13B is an A* algorithm based on the road network, which is embedded in the NetLogo network extension, `nw` (Wilensky, 1999). This algorithm applies to resident drivers.

For subway commuters, this chapter employs a Local Search Algorithm (LSA) for pathfinding (see Figure 5.14). LSA is an algorithm where the agent knows the goal state and the distance from the optimised path (termed error of distance) and asks the algorithm to rewrite the path to minimise further errors, which makes it memory efficient. A* was replaced with LSA because the algorithm that was asked to find the

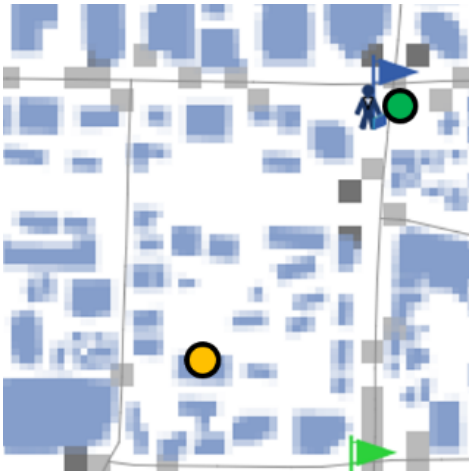


Fig. 5.14 The person next to the starting point (green) walks towards the goal point (yellow) following the shortest path, which is a straight line. Here, the agent decides to move closer to the goal point, but the route will be created at every step. Pedestrians penetrating the buildings is a downside of this method

lowest pollution patch between the current step and the final goal kept changing every step, which led to repetitive recalculation on every step, slowing the execution speed.

Amongst the searching functions of LSA, this study uses a "random-walk" or "hill-climbing search", where the agent iteratively searches the maximum value (or minimum value) within the boundary until it reaches the target. However, the function has a major drawback as the searching terminates either when it reaches the local maximum instead of the global maximum, or there is a huge plateau which does not have a higher surrounding value. Nevertheless, this study applied this method because the commuters in the CBD normally do not have any issues in getting lost when they are heading to work and back home. This study also created another scenario, *Awareness*, that asks agents to take an alternative route to avoid high PM_{10}. Details are documented in the scenario section.

Non-Exhaust Emissions and Dispersion

Recent studies from the UK and Europe equally documented the main sources of non-exhaust emissions such as tyre wear, brake wear, and road surface wear (Air Quality Expert Group, 2019; EMEP/EEA, 2019). A few papers included resuspension

as a fourth contributor, but this study articulates resuspension in the dispersion section below. Figure 5.15 illustrates the non-exhaust emissions, dispersion, and dilution.

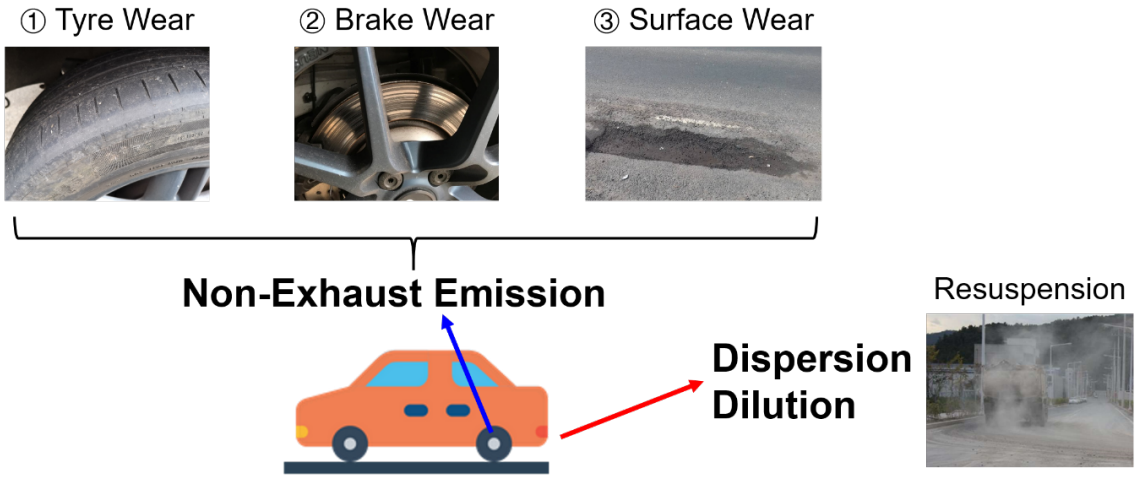


Fig. 5.15 Graphical explanation of non-exhaust emissions, dispersion, and dilution

According to EMEP/EEA (2019), the total of non-exhaust emissions is estimated with the following equation.

$$E_{total} = E_{Tyre} + E_{Brake} + E_{Road} \tag{5.2}$$

- E_{Total}: the total non-exhaust PM emissions
- E_{Tyre}: PM emissions from tyre wear
- E_{Brake}: PM emissions from brake wear
- E_{Road}: PM emissions due to road abrasion

Each component will be investigated in the following sections.

1) Tyre Wear

$$E_{Tyre} = \sum_{i=1}^{n} N_j \times M_j \times EF_{Tyre,j} \times F_{s,i} \times S(V) \tag{5.3}$$

- E_{Tyre}: Total emission for the defined time and spatial boundary (g/km)
- N_j: Number of vehicles in category j within the defined spatial boundary

Table 5.9 TSP (Total Suspended Particles) emission factors for source category road vehicle tyre wear (EMEP/EEA, 2019)

Vehicle class (j)	TSP emission factor (g/km)	Uncertainty range
Two-wheel vehicles	0.0046	0.0042 - 0.0053
Passenger cars	0.0107	0.0067 - 0.0162
Light-duty trucks	0.0169	0.0088 - 0.0217
Heavy-duty vehicles	Separate Equation	0.0227 - 0.0898

- M_j: Mileage (km) driven by each vehicle in category j during the defined time (not used)
- $EF_{Tyre,\ j}$: TSP mass emission factor for vehicles in category j (g/km)
- $F_{s,i}$: mass fraction of Particles that can be attributed to particle size class i
- S(V): Correction factor for a mean vehicle travelling speed V

As this equation was designed to measure the bulk emissions from a number of vehicles (e.g. 20g/km from 10 vehicles in a 5km trip between 10:00-15:00), it is not appropriate to measure the emissions of hundreds of vehicles that have separate journeys. To find a solution, this study manipulates N_j at an appropriate number based on sensitivity analysis, converts emission levels from g/km to µg/30m (equal to a size of one patch in the simulation), and spatial and temporal units at 30m and on a minute by minute basis.

For example, one passenger car (j) has an emission factor of .0107 (.0067-.0162) (g/km) (see Table 5.11), and to get an estimate of PM_{10}, the size distribution $F_{s,i}$ converts the TSP estimate to PM_{10} multiplying by a fraction of 0.6 (see Table 5.11). This can result in 32.1µg/m^3 per patch with an uncertainty range of 20.1 - 48.6.

In terms of vehicle speed, EEA sets the parameter V at 1.39 below 40km/h, and declining effect of (-0.00974 * V + 1.78) between 40-90km/h. It assumes that frequent brakes and accelerations are expected below 40km/h but less as the vehicle speeds up.

Table 5.10 Size distribution of tyre wear particles (EMEP/EEA, 2019)

Particle size class (i)	Mass Fraction of TSP
TSP	1
PM_{10}	0.6
$PM_{2.5}$	0.42
PM_1	0.06
$PM_{0.1}$	0.048

Table 5.11 Speed Correction (EMEP/EEA, 2019)

Velocity (km/h)	Factors (V)
V <40	1.39
40 $\leq$ V $\leq$ 90	-0.00974 * V + 1.78
V >90	0.902

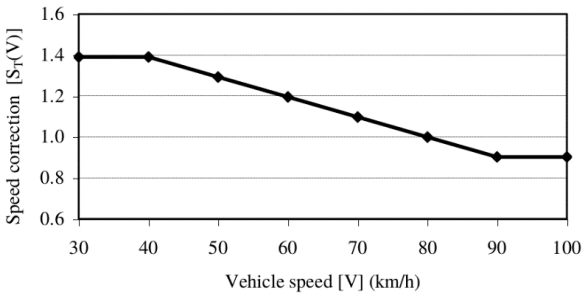


Fig. 5.16 Speed: tyre wear

Table 5.12 TSP (Total Suspended Particles) emission factors for source category road vehicle brake wear (EMEP/EEA, 2019). Here, this study only considers passenger cars.

Vehicle class (j)	TSP emission factor (g/km)	Uncertainty range
Two-wheeled vehicles	0.0037	0.0022 - 0.0050
Passenger cars	0.0075	0.0044 - 0.0010
Light-duty trucks	0.0117	0.0088 - 0.0145
Heavy-duty vehicles	Separate equation	0.0235 - 0.0420

2) Tyre Wear

The equation for brake wear is the same as tyre wear, and has only a few differences in parameters.

$$E_{Brake} = \sum_{i=1}^{n} N_j \times M_j \times EF_{Brake,j} \times F_{s,i} \times S(V) \tag{5.4}$$

- E_{Brake}: Total emission for the defined time and spatial boundary (g/km)
- N_j: Number of vehicles in category j within the defined spatial boundary
- M_j: Mileage (km) driven by each vehicle in category j during the defined time (not used)
- $EF_{Br,\,j}$: TSP mass emission factor from road wear for vehicles in category j (g/km)
- $F_{s,i}$: mass fraction of Particles that can be attributed to particle size class i
- S(V): Correction factor for a mean vehicle travelling speed V

As mentioned in the Tyre Wear section, emission factors for passenger cars must fit a unit set in the virtual environment. Thus, the $EF_{Br,j}$ value of .0075 (g/km) converts to 21.5 (µg/patch). The size distribution of PM_{10} is 0.98. The brake wear, particularly from the linings, are worn out quickly when the driver accelerates and decelerates frequently, and this tends to happen when the traffic volume is high.

Table 5.13 Size distribution of brake wear particles (EMEP/EEA, 2019)

Particle size class (i)	Mass fraction of TSP
TSP	1
PM_{10}	0.98
$PM_{2.5}$	0.39
PM_1	0.1
$PM_{0.1}$	0.08

Table 5.14 Speed Correction (EMEP/EEA, 2019)

Velocity (km/h)	Factors (V)
V <40	1.67
40 $\leq$ V $\leq$ 90	-0.0270 * V + 2.75
V >90	0.185

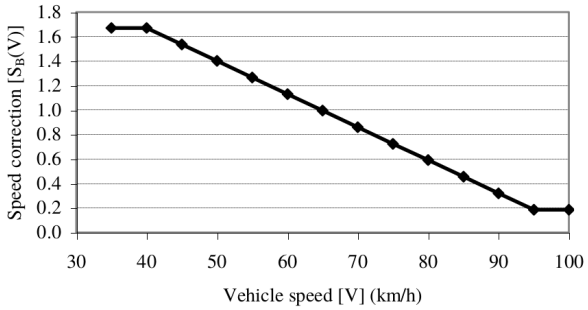


Fig. 5.17 Speed: Brake wear

Table 5.15 TSP emission factors from road surface wear (EMEP/EEA, 2019). Here, this study only considers passenger cars.

Vehicle class (j)	TSP emission factor (g/km)
Two-wheeled vehicles	0.006
Passenger cars	0.015
Light-duty trucks	0.015
Heavy-duty vehicles	0.076

Table 5.16 Size distribution of road surface wear particles (EMEP/EEA, 2019)

Particle size class (i)	Mass fraction (FR,i) of TSP
TSP	1
PM_{10}	0.5
$PM_{2.5}$	0.27

3) Surface Wear (i.e. Road Abrasion)

Road surface wear is caused by the appearance of wheel marks when the vehicle passes over the road or parts of the road are destroyed by heavy vehicles. The formula is as follows.

$$E_{Surface} = \sum_{i=1}^{n} N_j \times M_j \times EF_{SW,j} \tag{5.5}$$

- $E_{Surface}$: Total emissions for the defined time and spatial boundary (g/km)
- N_j: Number of vehicles in category j within the defined spatial boundary
- M_j: Mileage (km) driven by each vehicle in category j during the defined time (not used)
- $EF_{SW,j}$ = TSP mass emission factor from surface wear for vehicles in category j (g/km)
- $F_{s,i}$ = Mass fraction of TSP that can be attributed to particle size class i

Table 5.17 Conversion of Vehicle Speed in NetLogo

Original	Simulation
5km/h	0.25
10km/h	0.5
20km/h	1
40km/h	2

4) Dispersion and Dilution

There are many dispersion models applicable for exhaust emissions, but according to early research (Beevers et al., 2013; Panko et al., 2013), many things related to non-exhaust dispersion remain unknown. The University of California, Riverside (UCR) team is conducting an on-going project to understand the severity of non-exhaust emissions at nearer roads and is currently testing non-exhaust parameters in their existing dispersion model[11]. In line with the UCR project, this study also attempts to disperse pollution with a spread function, in-cone in NetLogo, as a surrogate of dust resuspension.

Dilution with non-combustible dust varies by meteorological or ventilation conditions. Less road dust would be generated on rainy days due to the additional weight that is deposited by the particle substances on the ground, and during night hours when there is less traffic. Cities like Seoul have employed water spraying trucks to spray moisture on the roads on dry days, which adheres the particles on the ground as well as keeps the resuspension low as possible. Since this study does not consider humidity or rain effects, the model will use the case from Nikolova et al. (2014), where it takes 110 seconds to dilute completely. In NetLogo, this is assigned as three random ticks – ranging between 0-2 minutes. This study further investigates the sensitivity of road PM_{10} by controlling both dispersion ranges and the extension of dilution.

5) Application Inside the Simulation

It is worth mentioning that the units change inside the *in silico* environment. Since one patch is equivalent to 30 metres and one car represents 10 vehicles, a car moving from one patch to the next means 10 cars moving 30 metres. The vehicle speed inside the simulation is assigned in Table 5.17.

[11] https://ww2.arb.ca.gov/resources/documents/brake-tire-wear-emissions

In published studies, the emissions are calculated by g/km based on the total distance of which the car has travelled (Ferm and Sjöberg, 2015; Srimuruganandam and Shiva Nagendra, 2010). Smit et al. (2010) argued that the atmospheric pollution is combined with emissions, humidity, wind, temperature, and other uncertain factors, and therefore the calibration process is normally tested in places where there are fewer confounding variables, e.g. tunnels. Calibration with observational values can be inaccurate, but more than 15 studies have chosen this method due to restricted conditions (Smit et al., 2010).

For example, if a car travels over a patch, it releases 10µg of tyre wear, 7µg of brake wear, 10µg of surface wear, and 3µg of resuspension. It will also have a dilution at 5µg. Thus, the total PM_{10} concentration would be the background PM_{10} + 25µg (Tyre + Brake + Surface + Resuspension – Dilution).

Health Loss and Recovery

The agent's health will decline on the assumption that it encounters over 100µg/m^3 at which they are currently located.

$$If PM_{10} \geq 100, \qquad \frac{dH}{dT} = -\alpha(H_{\max} - H(t)) + H_{\text{recov}} \tag{5.6}$$

While the equation above is equivalent to that of the previous chapter, there are several measurements in which the application differs from the previous chapter. First, the infiltration ratio, often termed as the I/O ratio, is used to estimate indoor exposure of individual agents. Infiltration ratio is applied to studies when only one has information about outdoor air pollution but less about indoor air pollution. Although the numbers seem quite simple, the ratio results from the consideration of the air exchange rate, windows opening, and type of housing. A few studies that used the I/O ratio also indicate that the ratio can vary by season (winter, summer) or types of microenvironments (classroom, house, office). This study chose the ratio from two studies, Kreider et al. (2019) where these authors took into account the I/O ratio from non-exhaust emissions, and Leung (2015) who reviewed a wide variety of households to get a parameter (see Table 5.18). The outdoor PM_{10} is assigned at 1, transit at 0.7, and indoors (including house and office spaces) at 0.2-0.7.

Table 5.18 Indoor-outdoor ratio of ambient PM_{10}

Type	Ratio
Outdoor	1
Transit	0.7
Indoors (house, office)	0.2-0.7

With the equation and infiltration ratio, the health loss for both subway commuters and resident drivers is applied under the same conditions. However, the difference would be their mode of transport and behaviours during weekends. Subway commuters spend their time exposed to ambient air pollution between the subway station and office. It is also considered that when the commuters travel out of the study area, they take more than an hour to arrive home[12]. Assuming the commuters stay at home between 11pm-6am, the commuter will be exposed to 0.2 times the ambient PM_{10} of the given patch. Resident drivers are mostly exposed to 0.7 times the ambient PM_{10} of the patch in transits and 0.2-0.7 times of that of PM_{10} when the vehicle is parked at the house or office. The vehicles are expected to be frequently exposed to high PM_{10} due to the substantial load of PM_{10} generated by road traffic.

Health recovery activates when the agent's health is below 100 and the agent is located at an indoor space. For a subway commuter, this will be when they are at home or office, while drivers recover when the car is parked. The recovery rate is given an arbitrary number of 10 by each minute but stops working when the health of an individual goes above 100.

Scenario Forecasting

This section outlines how vehicle prohibition can effectively improve air quality in Seoul CBD, as well as how people's information and awareness can prevent exposure to air pollution. The scenario was designed based on the 'Green Transport Scheme' initiated in December 2019, and thus it attempts to help measure the effectiveness of implementation that is already in place.

The Green Transport Scheme aims to improve air quality in Seoul by restricting high-emission vehicles from entering the CBD area. The municipal government restricts

[12]Joong-ang daily article, March 7th 2019, "South Korea's office workers spend 103 minutes on average to get to work"

Grade 5 vehicles, mostly diesel cars, between 06:00-21:00, and violators are fined 100USD. This study looks at how the effects of non-exhaust emissions resulted from barring vehicle entry and illustrated how people's health might improve from the scenario.

The first scenario is to restrict extra inbound vehicles. It measures how PM_{10} will improve if vehicles are restricted by 50% or 90%. **The second scenario compares the outcome of the population at risk depending on the awareness of individuals to extreme PM_{10}**. When the awareness scenario is not activated, the subway commuters will walk on the shortest distance to their destination and the resident drivers will take free trips within or outside the CBD over the weekend regardless of their health. When the awareness scenario is activated, the subway commuters either walk on the path that does not exceed 100µg/m^3 of PM_{10} or on the lowest value of three patches in front of their path when all the surrounding patches exceed 100µg/m^3. The drivers below the nominal health of 100 will not take a journey. Both scenarios are implemented in combination.

5.5 Sensitivity analysis and Calibration

This section experiments with the sensitivity of selected parameters and calibrates the modelled pollution outcomes with PM_{10} observations. Each parameter was analysed from an average of 20 iterations that reduced possible stochastic effects.

Local sensitivity models tend to be one-factor-at-a-time (OFAT), where an input factor changes its quantity while the other input factor has their values fixed. OFAT is useful to test the sensitivity between the main effects at the first step. However, one should be aware that this method ignores the interaction between factors which can result in a different result if another parameter value was chosen (Thiele et al., 2014). Thus, normally, studies tend to use global sensitivity methods having tested the sensitivity for each parameter locally.

Despite the advantages seen in global sensitivity, this study used the one-factor-at-a-time (OFAT) method to examine the sensitivity for each of the parameters. The main reason was that there was a memory ceiling which was not sufficient to consider five parameters over 130,000 minutes (the model stopped executing after a two-week run). Having tested two parameters over a selected period (i.e. days when the background PM_{10} exceeded 100µg/m^3), there were no noticeable interaction effects discovered in the outcome.

5.5.1 Sensitivity: One-factor-at-a-time (OFAT)

PM_{10} levels by Emission Factors

Table 5.19 and Figure 5.18 show how PM_{10} results from the emission factors of vehicle agents. As a reminder, the emission parameter is the number of cars (N) represented in the equations of Section 5.4.7. It can be interpreted as how many cars have polluted PM_{10} on this patch.

It turned out that the levels of PM_{10} increased linearly as the emission factors increased. Table 5.19 indicated that the mean PM_{10} of Jongno was 43.4µg/m^3, 60µg/m^3, 81.4µg/m^3, and 123µg/m^3 in emission factors 1, 5, 10, and 20 respectively. The difference between each factor was 16.6µg/m^3, 21.4µg/m^3, and 41.6µg/m^3, which increased proportionally as the factors increased. This linear increase was not only was seen in the mean figure but also seen on any of the dates, including the peak value

on January the 20th where the levels sat near 150µg/m^3 in factor 5, but showed an increase to around 200µg/m^3 and 250µg/m^3 from factors 10 and 20 (see details in Figure 5.18).

PM_{10} between roads varied greatly when the emission parameter was high. Although the model did not give any direction to the vehicles nor the hierarchy of roads, PM_{10} levels varied by 12µg/mm^3 in Emission 20, where the lowest was 122.6µg/mm^3 at Sejong and the highest was 134.1µg/mm^3 at Yulgok (see Table 5.19). This implies that although the number of road lanes was not specified, the high parameter value can measure the variability of PM_{10} by roads.

Table 5.19 PM_{10} concentrations in five CBD roads based on emission factors of 1, 5, 10, and 20

Date	Emission factor	Jongno	Sejong	Yulgok	Samil	Pirun
Overall	1	43.4	43.2	42.8	42.8	42.9
	5	60.0	60.1	62.0	61.6	61.7
	10	81.4	81.2	85.6	85.3	85.2
	20	123.3	122.6	134.1	132.7	133.4
Jan 8th	1	49.2	48.6	48.1	47.8	48.4
	5	65.5	66.7	67.8	67.9	66.9
	10	85.3	86.4	90.3	91.9	90.5
	20	129.6	115.3	141.1	150.1	128.7
Jan 15th	1	58.1	57.6	57.1	56.9	57.5
	5	75.0	75.2	77.0	76.9	75.1
	10	93.8	104.3	100.8	101.6	97.1
	20	136.1	133.3	147.8	154.4	148.6
Jan 22nd	1	37.7	37.4	37.1	37.0	37.2
	5	53.9	52.5	55.1	57.3	55.4
	10	75.9	78.4	78.5	85.0	76.9
	20	114.6	117.6	127.6	127.6	124.3

Dispersion and Dilution

This section examines the sensitivity of dispersion and dilution parameters that affect roadside PM_{10}. The variables are conceptualised in Figure 5.19. By default, each vehicle disperses non-exhaust PM_{10} onto the neighbouring patches by an angle of 60° which dilute in 0-3 minutes. Having controlled the duration of dilution (< 3 ticks), the first experiment simulated the range of dispersion at 45° and 90°. Then, controlling

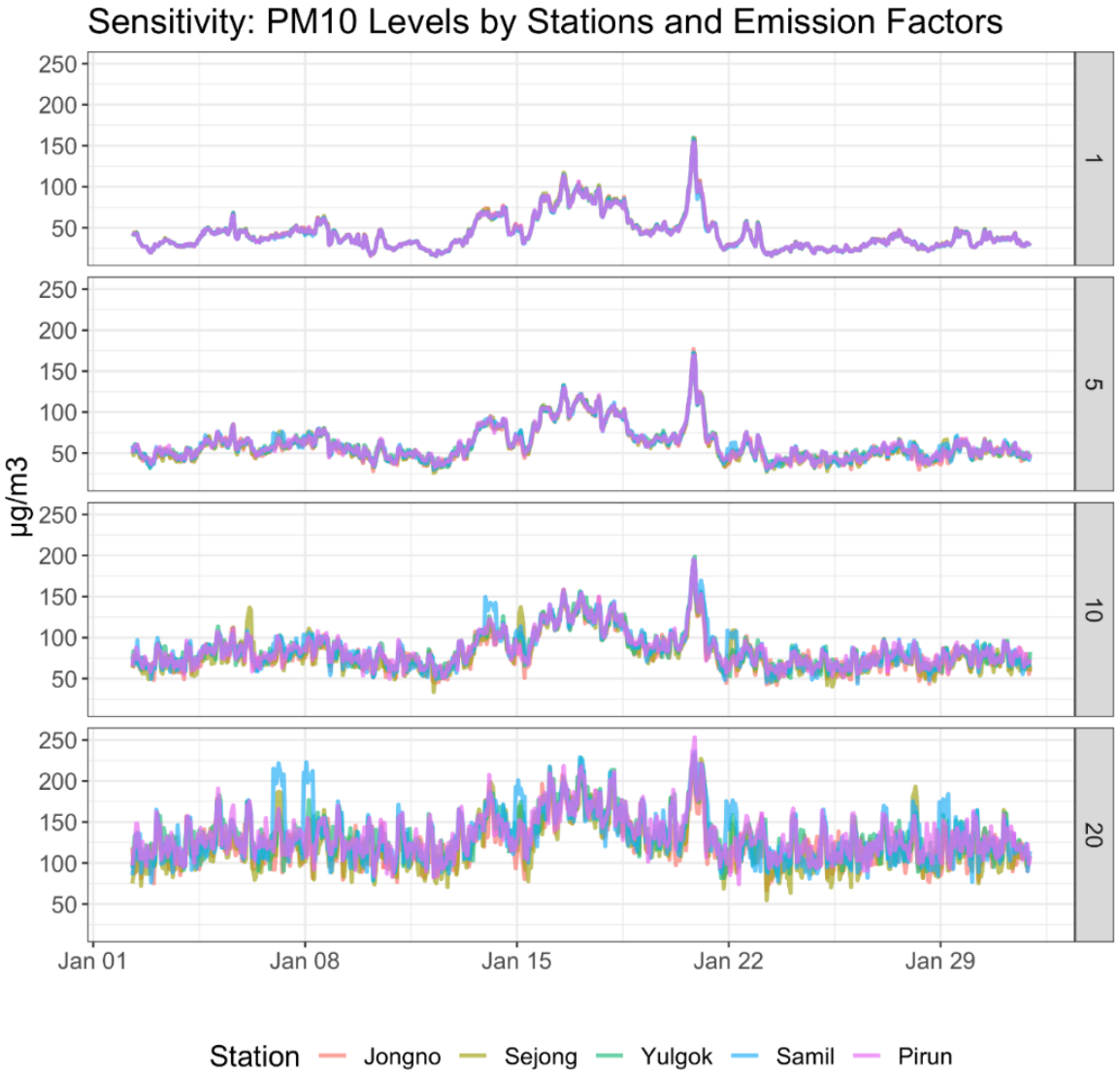


Fig. 5.18 PM_{10} levels by emission factors of 1, 5, 10, and 20, each showing the N of vehicles that generate non-exhaust PM_{10} emission. Each line represents 5 sample points of Jongno, Sejong, Yulgok, Samil, and Pirun roads. The variability at any station increases as the emission factor is increased

the dispersion to 60°, the next experiment simulated the dilution process by $5+\alpha$ and $10+\alpha$.

Results showed that dispersion range displayed less sensitivity on roadside PM_{10}, except for Jongno, where the difference of cone width between 45° and 90° was around $3\mu g/m^3$ in `Emission 5` and `Emission 10`, and further increased to $14\mu g/m^3$ in Emission

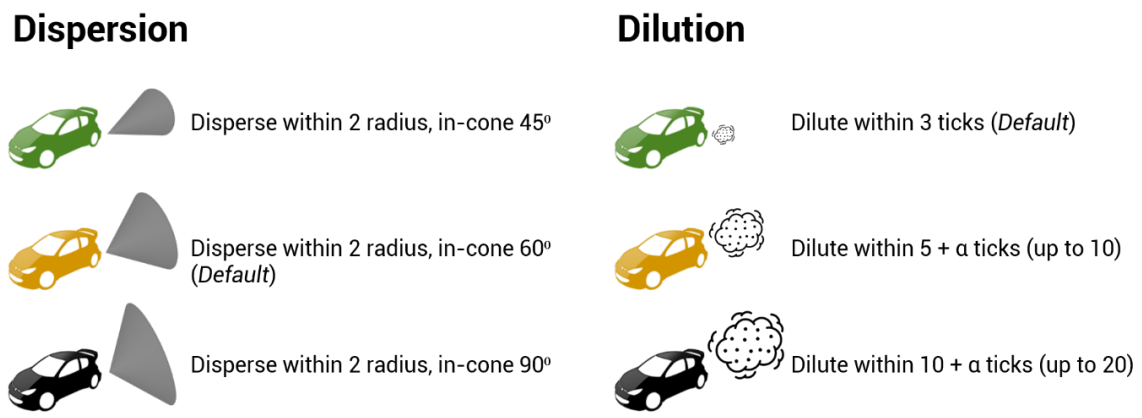


Fig. 5.19 Illustrations of dispersion parameters (left) and dilution parameters (right)

20 (see Table 5.20). This implies that the range of dispersion might not be sensitive to the PM_{10} on-roads, such as Sejong and Pirun stations, but from the evidence of Jongno, a distant station, it may deliver higher PM_{10} to people walking near roads.

Table 5.20 PM_{10} concentrations by emission factors and dispersion range (cone width) (Unit: µg/m^3)

Emission	Cone Width (°)	Jongno	Sejong	Yulgok	Samil	Pirun
	45	50	49.6	50.5	50.3	50.9
1	60	50.3	49.8	50.5	50.4	51.1
	90	50.7	49.9	50.6	50.5	51.1
	45	58.4	55.7	58.7	58.9	60.1
5	60	59.3	56.3	59	59.5	60.4
	90	60.4	56.6	59.1	59.5	60.8
	45	73.2	71.2	77.3	77	80.5
10	60	76.6	72.3	77.9	77.4	81
	90	79.6	73.1	78.5	78.1	81.8
	45	102	100	112	113	118
20	60	109	102	114	115	120
	90	116	104	115	118	120

Unlike dispersion results, all roads were very sensitive to the dilution period except for Emission 1 (see Table 5.21). In `Emission 5`, the default period of less than 3 minutes indicated an average figure of 60-62µg/m^3, however, extending the period to 10 minutes increased PM_{10} to 67-69µg/m^3, which was 10% higher than the default.

The difference between dilution periods increased proportionately to emission factors where the quickest (3 mins) was 14-18µg/m^3 higher than the slowest (10 mins) in `Emission 10` and 31-41µg/m^3 in `Emission 20`. If this analysis was to represent the length of dust resuspension in the real world, say 3 minutes of dust floating until dilution, the deterioration of PM_{10} can be explained by the floating particles from the vehicles that mixed well with the atmosphere. A disclaimer is that the dilution is only affected by the duration of ticks (zero wind), and no other components (e.g. wind, rain) that change dilution time.

Table 5.21 PM_{10} concentrations by emission factors and (the duration until) dilution (Unit: µg/m^3)

Emission	Duration	Jongno	Sejong	Yulgok	Samil	Pirun
	3	45.5	45.8	45.8	46	46.1
1	5	46.1	46	46.5	46.2	46.4
	10	46.7	46.5	46.8	46.7	47
	3	60	60	62	62	62
5	5	66	66	66	66	67
	10	67	67	68	68	69
	3	81	81	86	85	85
10	5	94	95	96	96	99
	10	99	99	100	100	102
	3	123	123	134	133	133
20	5	153	150	155	155	159
	10	164	160	165	164 167	

PM_{10} Levels by Car Ratio

This section investigated how PM_{10} can be sensitive to changes in car sampling (see Figure 5.20). Resident vehicles were not included in this experiment as short-term journeys from the resident vehicles hardly contributed emission levels to the result (these were tested but not included in the thesis). To summarise, car ratios of 0%, 2.5%, 5%, 10%, and 20% mean sample rates of traffic counts by each minute were taken from the traffic monitoring statistics.

In the 0% run, assuming no other vehicles, the roadside PM_{10} ranged between 47-50µg/m^3 (which is equal to the background level), which was at least 10µg/m^3 lower than the concentrations from other ratios. However, different sample sizes merely

showed a small difference. For example, a 10% sample in Jongno only contributed 1.8µg/m^3 more than that of 2.5%.

Surprisingly, all roads showed less pollution in the 20% sample because a massive number of vehicles failed to enter the study area. The queues were particularly long in Samil and Yulgok roads because Samil had fewer traffic signals at the entrance of the road which enabled vehicles to accelerate up to the core area with a few ticks but soon met several junctions, which can be depicted as a bottleneck effect; Yulgok has a roundabout that reduces the speed.

Table 5.22 Car Ratio and PM_{10} concentration (Unit: µg/m^3)

Ratio	Jongno	Sejong	Yulgok	Samil	Pirun
0%	50.5	49.5	48.0	50.2	47.8
2.5%	59.5	60.0	60.3	60.9	61.4
5%	61.9	61.6	62.7	64.2	64.2
10%	62.3	61.8	63.4	64.4	64.4
20%	60.8	61.5	63.6	64.0	64.0

Health Loss

This section investigated the health risk of subway commuters and resident drivers who are sensitive to the health loss parameters. Here, individuals only lose health when PM_{10} exceeds 100µg/m^3, and contribute to the population at risk when one's health status falls below 100. Output A of each figure resulted from allowing extra inbound traffic in the CBD, whereas output B of each figure resulted from no other traffic than the resident vehicles.

For subway commuters, the population at risk appeared on January 20th-22nd, late February, early March, and late March (see Figure 5.21A). The maximum risk rate was 10% in `0.03` and proportionately increased to 30% in `0.1`, but suddenly skyrocketed to 100% over `0.15`. Although a lot of uncertainty from other parameters has contributed towards the outcomes, the tipping point of the health loss parameter was somewhere between `0.1` and `0.15`. Several oscillations were also discovered during the extreme PM_{10} events. This was because subway commuters have different commute hours that led them to be exposed to ambient PM_{10}, and since health recovery activates when the individual arrives at home or the workplace, the risk rate oscillates frequently.

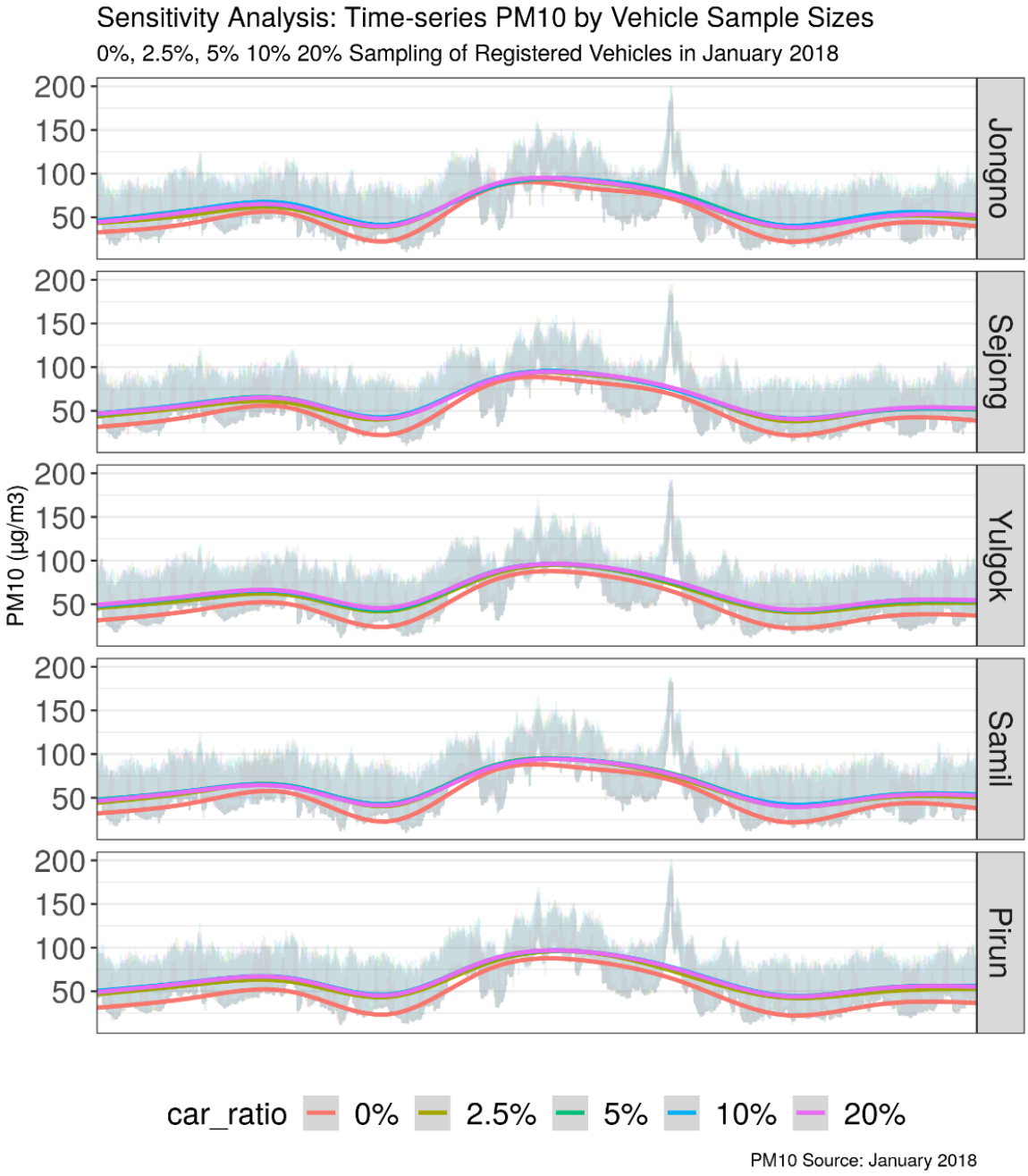


Fig. 5.20 PM_{10} levels by car ratios of 0, 2.5%, 5%, 10%, and 20%. The average values (smooth curve) of PM_{10} is similar across roads but was different between 0% (no extra cars in the CBD) and the rest of the samples.

With a car-free experiment (see Figure 5.21B), the results did not affect the health risks of subway commuters. This is because the trajectories of the commuters between stations and office locations were mostly distant from the road. However, the sensitivity

between health loss parameters was comparable to the previous experiment: health risk proportionately rose until the parameter reached `0.1` but a sudden upsurge appeared when the parameter was over `0.15`.

Compared to subway commuters, resident drivers experienced fewer occurrences of health risk, but higher surges in extreme PM_{10} episodes particularly over the parameter value of 0.15 (see Figure 5.22A). Throughout the whole period, the health risk of resident drivers emerged during January 22nd, February 12th, March 8th, and March 24-25th, where the majority was at risk at the last peak. The prominent difference by the health-loss adjustment was very clear at the first peak where it started from less than a percent of risk at `0.03`, then rose to 2.5% and 6% in `0.05` and `0.1`, then surged to 50% and 71% on `0.15` and `0.2`. In line with the subway commuters, a tipping point was also seen between `0.1` and `0.15`.

However, the first surge that happened in `0.15` and `0.2` experiments significantly reduced to 15% and 18% in a car-free condition (see Figure 5.22B). The other parameters only showed a less than 2% difference. This implies that the health risk of the drivers was not only sensitive to the health-loss parameters but also was affected by the emissions generated by non-resident traffic.

The difference in health risk can differ by the time the individual has spent outdoors when the ambient PM_{10} is over the threshold of 100µg/m^3, and how quickly that person recovered health. Even if 30% of subway commuters have experienced health risk, the short walking distance allowed them to recover promptly. By contrast, although drivers had fewer emergences of health risk, traffic congestion together with high background pollution had rapidly deteriorated the driver's health, especially on extremely polluted days.

This study chose one subway commuter and one driver to understand how the nominal health changed over time (see Figure 5.23A). The light shaded colours shown in the background is the health status by each minute and the lines of turquoise and red are the moving averages. The health status of a subway commuter lost health earlier than the driver under the same condition. The driver might seem healthier than the pedestrian because the driver was never exposed to ambient PM_{10} which prevented multiple threats of major PM_{10} episodes. In Figure 21B, the selected driver experienced fewer health risks in the car-free experiment, which can support the result of the population outcomes in Figure 5.22B where a major fall in risk rate is for drivers in a car-free situation.

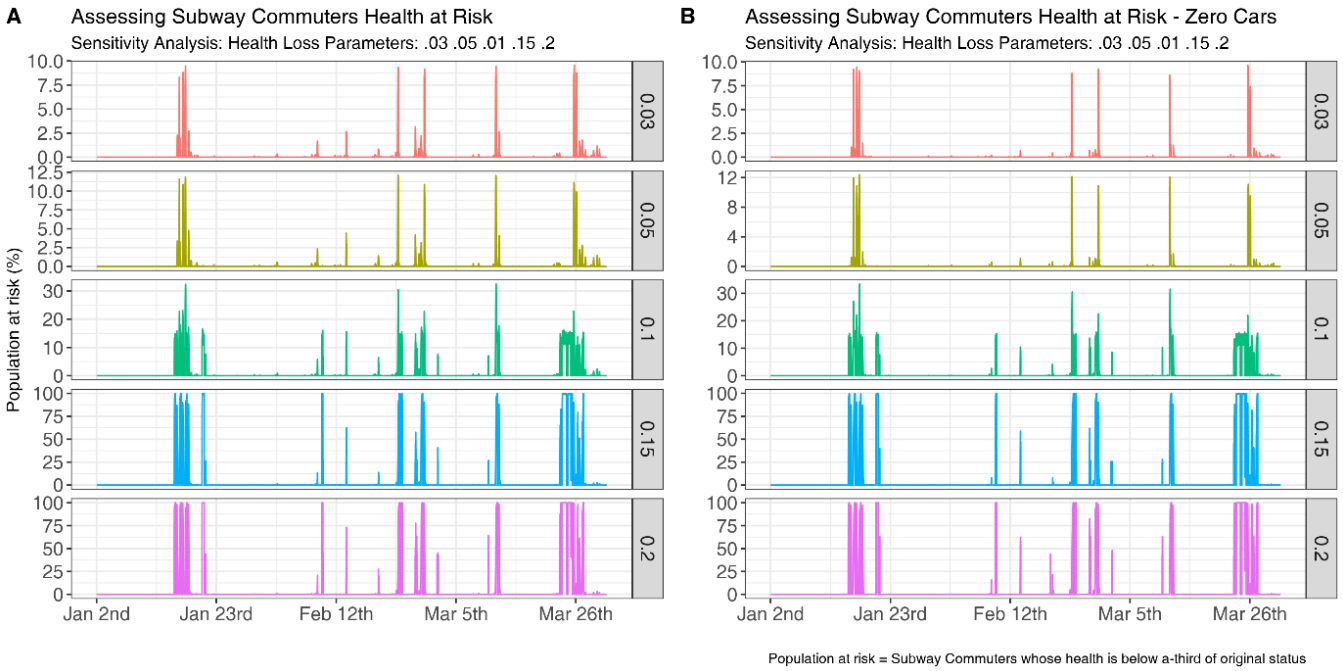


Fig. 5.21 Temporal change of risk rates for subway commuters (% of those with health under 100) with inbound cars (left), and without inbound cars (right)

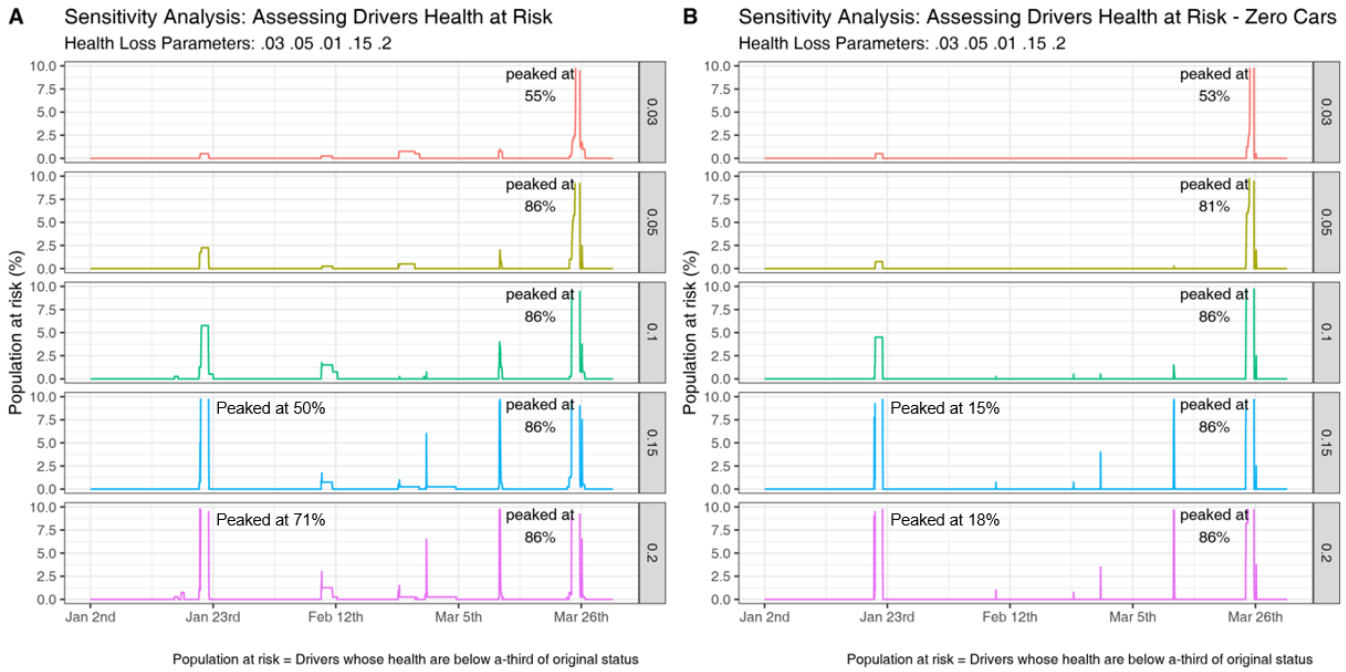


Fig. 5.22 Temporal change of risk rates for resident drivers (% of those with health under 100) with inbound cars (left), and without inbound cars (right)

The rolling mean between the two groups converged as the parameters increased (see Figure 5.23). The subway commuter's health was almost the same in different patterns, but for the drivers, the high parameter settings might have caused higher health loss even from a single pollution episode. The difference exists between the two on January 23rd because of the indoor factor of 0.7 that benefited the vehicle drivers.

In short, signs of deterioration in health appeared continuously in long-distance commuters on days when PM_{10} was on the rise, while the resident drivers had a relatively short period of commute time that prevented frequent health risk, but the extreme levels of PM_{10} led most of the drivers to an acute health crisis.

Walking Speed

To test how walking speed affects the change to risk population, this section adjusted various levels of walking speed for subway commuters. Given the default speed at 0.6-1.0, the section tested 1) 0.2-0.4 patch per minute, 2) 0.4-0.7 patch per minute, and 3) 1.6-1.8 patch per minute. The range was given under the assumption that people have different walking speeds. Walking speed over .5 might seem rather unrealistic, but this experiment intended to illustrate how speed affects exposure levels.

The time series graph clearly showed that the onset and peak levels were very sensitive to walking speed (see Figure 5.24). When the pedestrian's walking speed was "Extremely Slow" (0.2-0.4), more than 40% of the population was at risk on five different occasions with the highest peak of 47%. However, the risk rate declined by 10% when the walking speed increased to "Slow" (0.4-0.7) and further declined by 30% when the speed increased to 1.6-1.8. This corresponds to the previous sensitivity analyses because slowing down the walking speed can mean that the person is prolonging the exposure time, which in turn causes a further health loss.

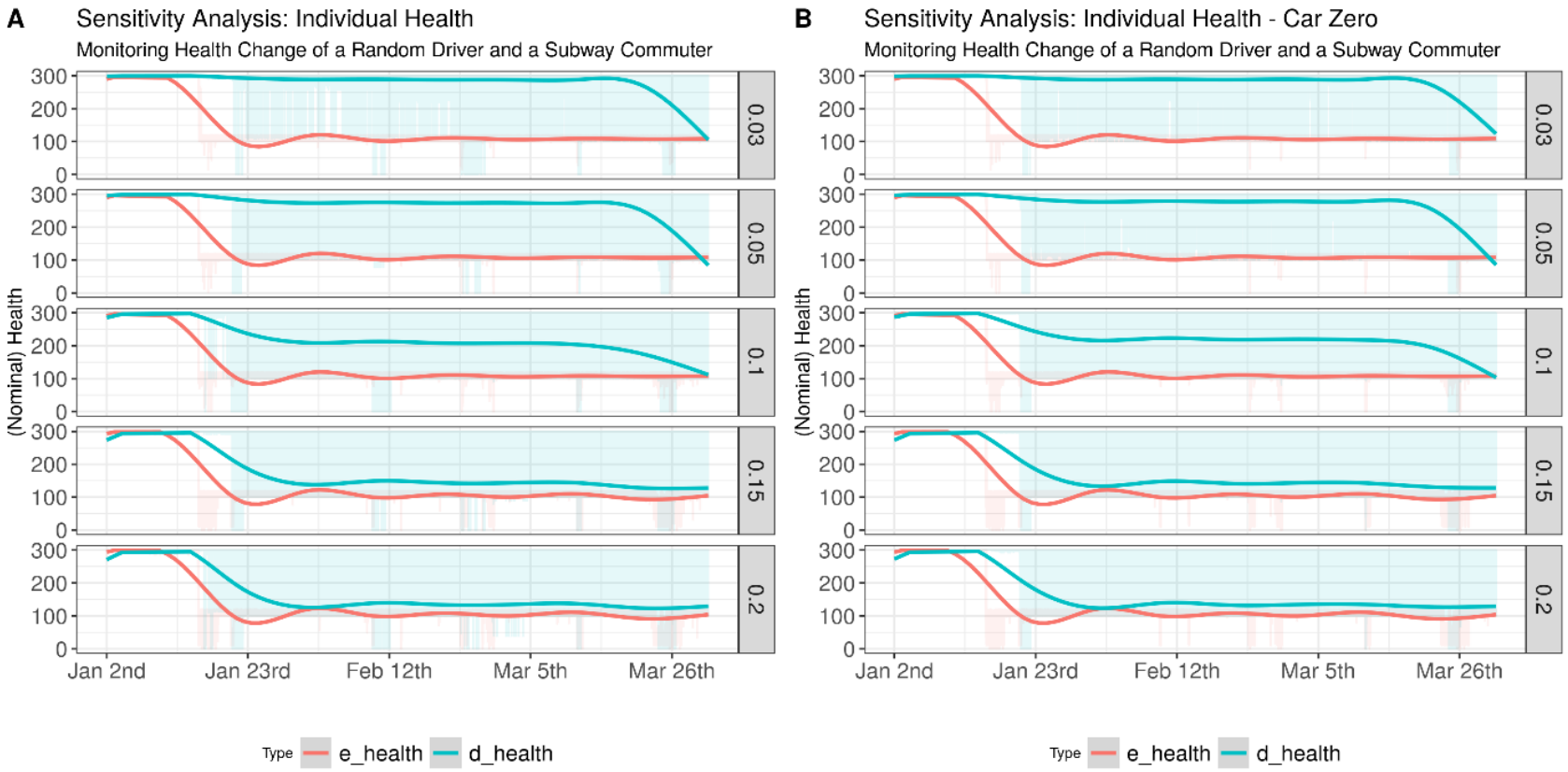


Fig. 5.23 Health comparison between a randomly chosen subway commuter (e_health) and a resident driver (d_health) with the case of (A)traffic and (B) traffic-free

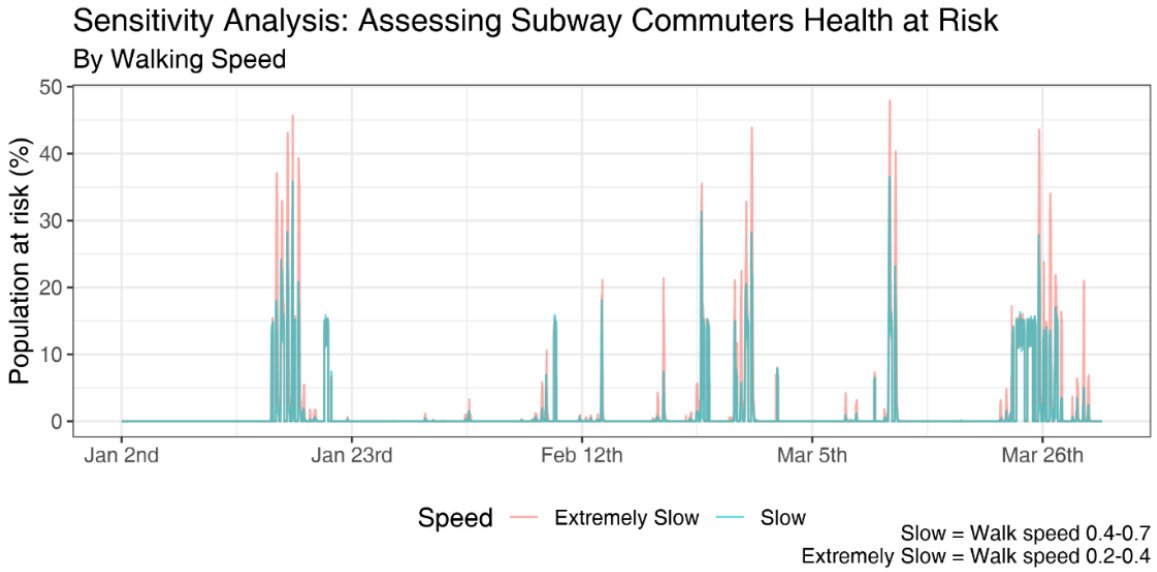


Fig. 5.24 Assessing subway commuters' health by different walking speed parameters

5.5.2 Spatial Output

Another sensitivity study was conducted using a spatial analysis to examine how PM_{10} concentration varies in space. PM_{10} on the 2nd of January was selected for this test, and minute-based results were aggregated to an hourly scale.

Time series pollution maps of an hourly mean and max were illustrated in Figure 5.25. Overall, the maximum roadside PM_{10} was at least two times higher compared to the background PM_{10}, where the road nearer to the junctions were over 100µg/m^3, but on average, the roadside PM_{10} was comparable to the background level, although the levels closer to the junctions were higher. Figure 5.25B (max. value) shows that all the roads had experienced over 80µg/m^3, which was distinctively higher than the background level, and the junctions in larger roads have frequently exceeded 100µg/m^3 during rush hours. It is speculated that the junctions produce more PM_{10} from the driver's stop-and-go behaviour. This is similar to the previous findings that brake emissions vary spatially and tend to escalate as it gets closer to the junction or down hill (Air Quality Expert Group, 2019; Timmers and Achten, 2018).

At 10:00am, the selected locations of Jongno and Samil had comparable levels of 50µg/m^3 on average and a maximum of 93µg/m^3 (see Figures 5.26 and 5.27). Although Jongno was slightly away from the junction, a large number of vehicles

had emitted particulates that increased and prolonged PM_{10} on road patches. Also, relatively higher PM_{10} was observed in areas where more traffic signals are installed or near the CBD entry points where a lot of vehicles attempt to enter simultaneously. This can mean that even on the same day, people who commute near the areas either by car or on foot have higher chances to be exposed to polluted air. Thus, more attention is needed when walking near the junctions where vehicles are likely to generate more particles from their tyres and brake wear.

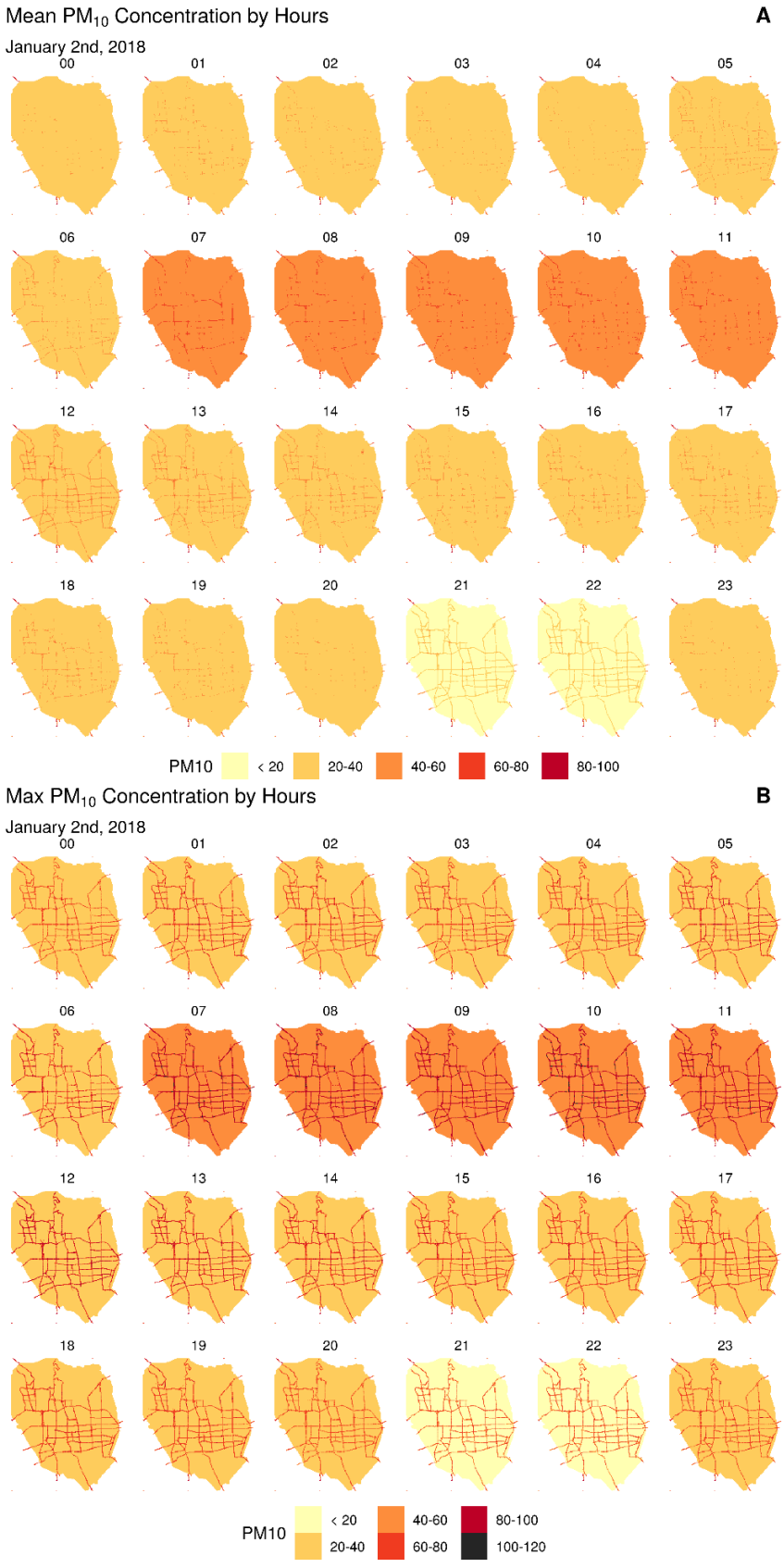


Fig. 5.25 Mean PM_{10} (top) and Max PM_{10}

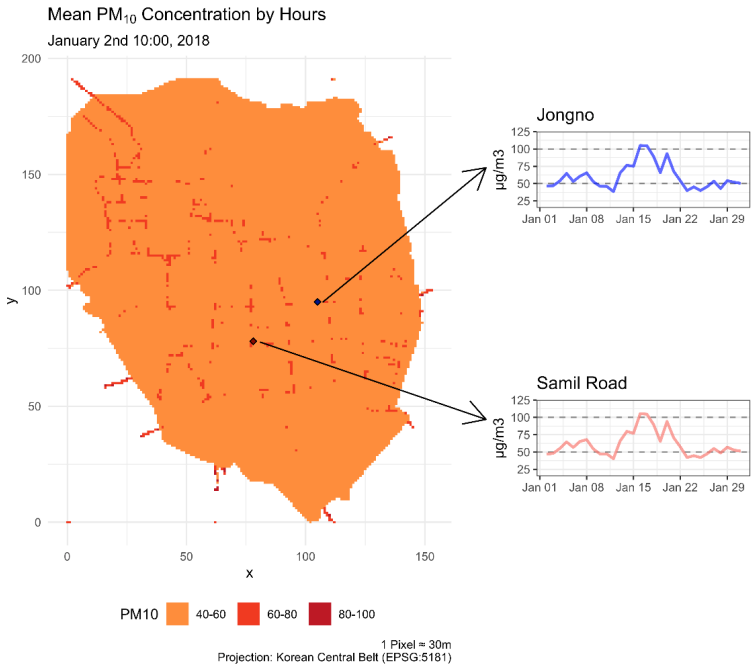


Fig. 5.26 Mean PM_{10} at 10:00

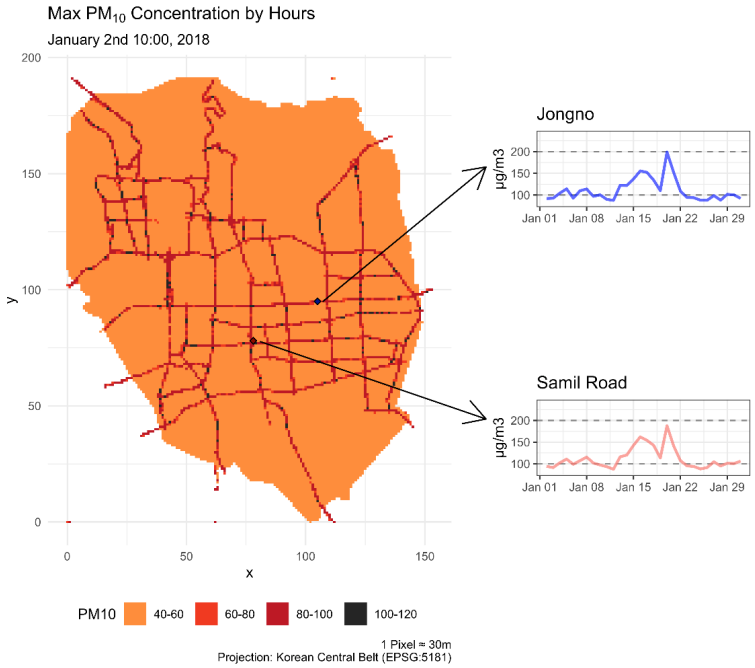


Fig. 5.27 Max PM_{10} at 10:00

5.5.3 Calibration

This section calibrates emission factors with the observation values measured from Jongno roadside station. From the sensitivity analysis, it was found that the emission parameters were not only sensitive to the increase in the parameter but also to the variation between roads when the parameter was over 10. Calibration in ABM is very common as it controls the errors and the uncertainty close to the acceptable level (Grimm et al., 2020).

This section did not calibrate PM_{10} across the study area because the background PM_{10}, which covers most areas, was already generated by the station data. Simulation results of Jongno were averaged from 20 iterations to avoid any noise from a particular run, then compared with the observation value. The simulation ran from the 2nd to the 31st of January 2018. This study used mean squared errors (MSE) and regression to examine the robustness of the model (see Figure 5.28). MSE, as it is known, as the average squared difference between the estimated values and the actual value, can be used to compare the results in positive numbers, understanding the values closer to zero are more accurate. R^2 is useful because it is often easier to interpret since it doesn't depend on the scale of the data, and people are familiar with percentages. Note here that each method has its pros and cons and there is no ground rule in selecting a method.

On January the 8th, MSE varied largely by 25, 42, 419, and 2684 in parameters 1, 5, 10, and 20; they varied by 90, 6, 248, and 2087 on the 15th, and 26, 42, 481, and 2426 on the 22nd. Throughout the whole month, `Emission 1` and `Emission 5` had the lowest MSE values by 18 days and 12 days respectively. The line graph shows that low biases for high values are observed in `Emission 1`, whereas high biases for low values are observed in the other parameters, but all modelled parameters could not replicate the peak values introduced from the observed values.

Regression results are similar to the MSE results, where the R^2 appeared to be highest in the lower two parameters and decreased significantly in the upper two parameters (see Figure 5.28B). In line with the MSE results, the scatter plot from `Emission 1` underestimated the observation values, in which most of the points were concentrated on the right side of the 1:1 line. `Emission 5`, on the other hand, slightly overestimated the results on the lower values but got most of the values, including the high values closer to the 1:1 line. Hence, although the overall MSE was lower and R^2 was

Table 5.23 Sum of standardized squared errors (SSSE) on January the 8th, 15th, and 22nd.

Date	Emission	Model	Observation	MSE
Jan 8th	1	49.2	56.4	25
	5	65.5	56.4	42
	10	85.3	56.4	419
	20	129.6	56.4	2684
Jan 15th	1	58.1	71.5	90
	5	75	71.5	6
	10	93.8	71.5	248
	20	136.1	71.5	2087
Jan 22nd	1	37.7	44.9	26
	5	53.9	44.9	41
	10	75.9	44.9	482
	20	114.6	44.9	2426

higher in `Emission 1`, the author selected `Emission 5` as the correct parameter. The reason being, that `Emission 5` effectively expresses the extreme values on a polluted day, while at the same time predicting closer values to the truth value. `Emission 1`, even on a minute-by-minute basis, does not articulate the peak of particulates that have possibly dispersed into the local atmosphere.

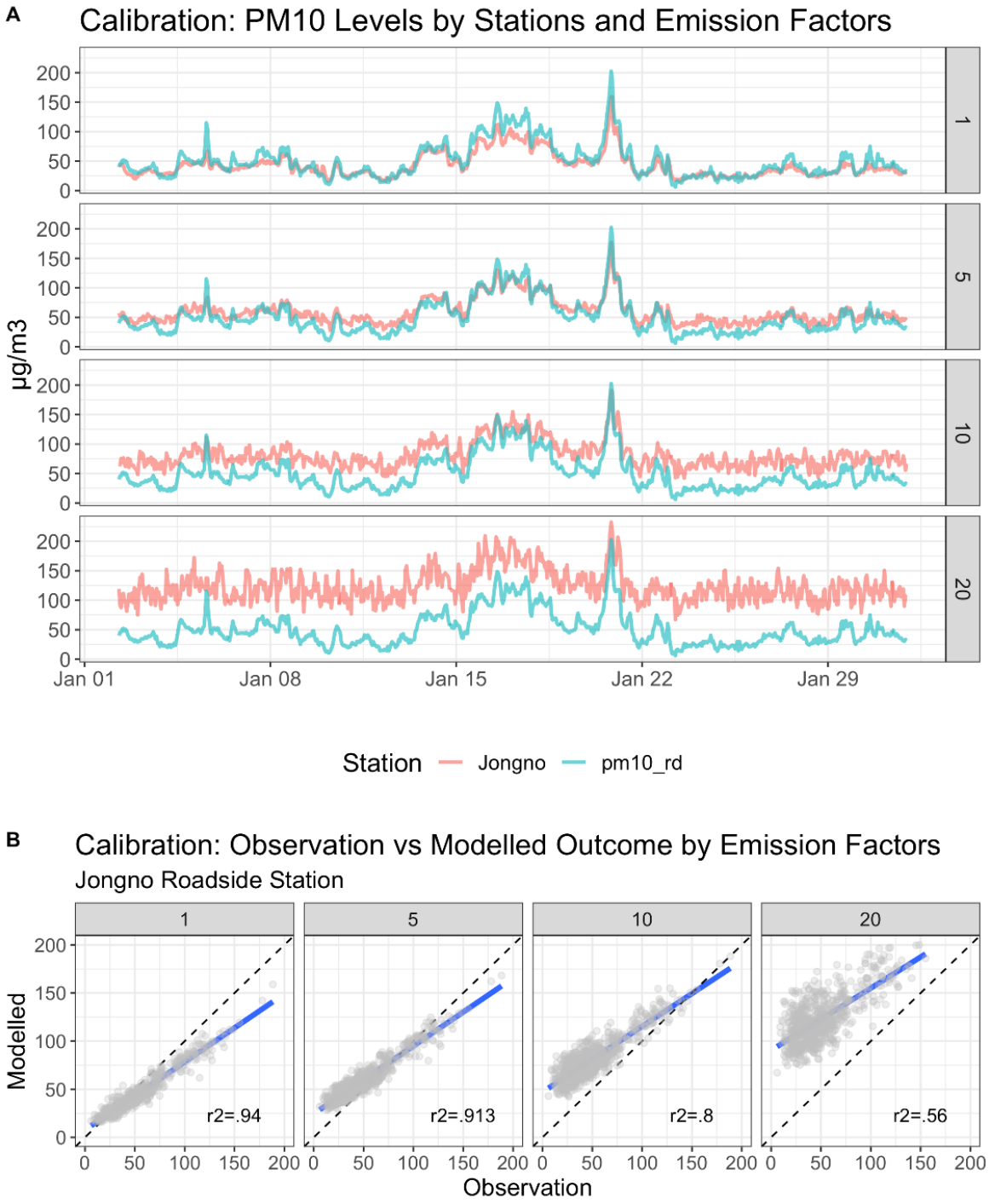


Fig. 5.28 Sensitivity output of adjusting the emission factor N from the equation of non-exhaust emissions (A) Jongno: Modelled, pm10_rd: Observation); (B) correlates the modelled output against the observation of Jongno roadside station. R2 of factors 1, 5, 10, and 20 returns .94, .91, .8, and .56.

5.6 Scenario Forecasting

Having calibrated the sensitivity of the parameters, this section conducted 'what-if' scenarios to understand 1) how vehicle restrictions can improve air quality, and 2) how the walking commuters can alter their paths when the pollution is high. As with the sensitivity section, the models were averaged from 20 iterations.

5.6.1 Vehicle Restriction

Based on different restriction scenarios of traffic entry, it turns out that vehicle restriction improved air quality (see Table 5.24). Compared to Business-as-Usual, restricting 50% of the inbound traffic reduced PM_{10} by 1.2-2.7µg/m^3 (2-4%) for the selected roads, while a 90% restriction scenario significantly reduced PM_{10} by 11.4-15.7µg/m^3 (18-24%). This corroborates the study where reducing brake disc emissions by 50% can decrease less ambient PM_{10} by 4-14% (Perricone et al., 2018).

The temporal trend of PM_{10} was comparable by roads as well as scenarios (see Figure5.29). Since the model did not take into account travel directions for non-resident vehicles nor the number of lanes in roads which can determine the routes of vehicles, the PM_{10} levels did not vary between roads.

Boxplot results also intuitively supported scenario results (see Figure 5.29). As with Table 5.24, the mean PM_{10} at 50% restriction did not show a significant change, but a noticeable decrease in PM_{10} was found in the 90% scenario where both the averages fell due to the extreme emission values that used to be generated from the traffic. This entails that a ban on vehicles has improved the air quality by decreasing the overall PM_{10} from 60µg/m^3 to 47µg/m^3, but more importantly, the change was due to the removal of extreme values, which may prevent instantaneous harm on human health.

Table 5.24 Overall average of PM_{10} on five roads by car restriction scenarios

Scenario	Type	Jongno	Sejong	Yulgok	Samil	Pirun
Business as Usual	PM_{10}	60.7	61.3	62.5	62.4	63.6
50% Restriction	PM_{10}	58.0	59.6	60.2	61.2	61.3
	Difference to BAU	2.7	1.7	2.3	1.2	2.3
90% Restriction	PM_{10}	47.9	49.8	48.1	51.0	47.9
	Difference to BAU	12.8	11.5	14.4	11.4	15.7

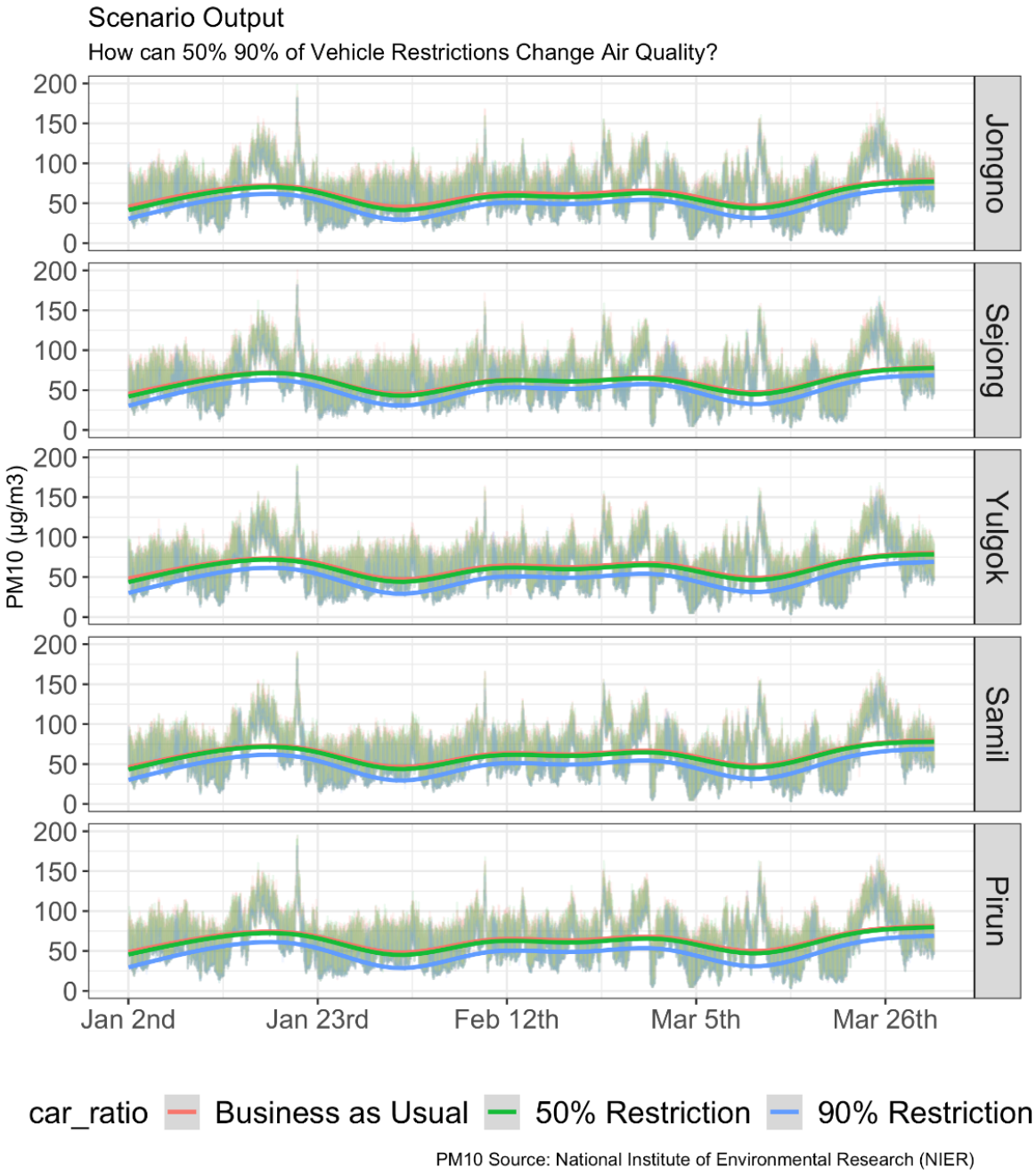


Fig. 5.29 A temporal trend of PM_{10} on five roads by car restriction scenarios

5.6.2 People's Awareness

In terms of assuming people are aware of their local air pollution boundaries or not, the risk to the population varied substantially between subway commuters and resident

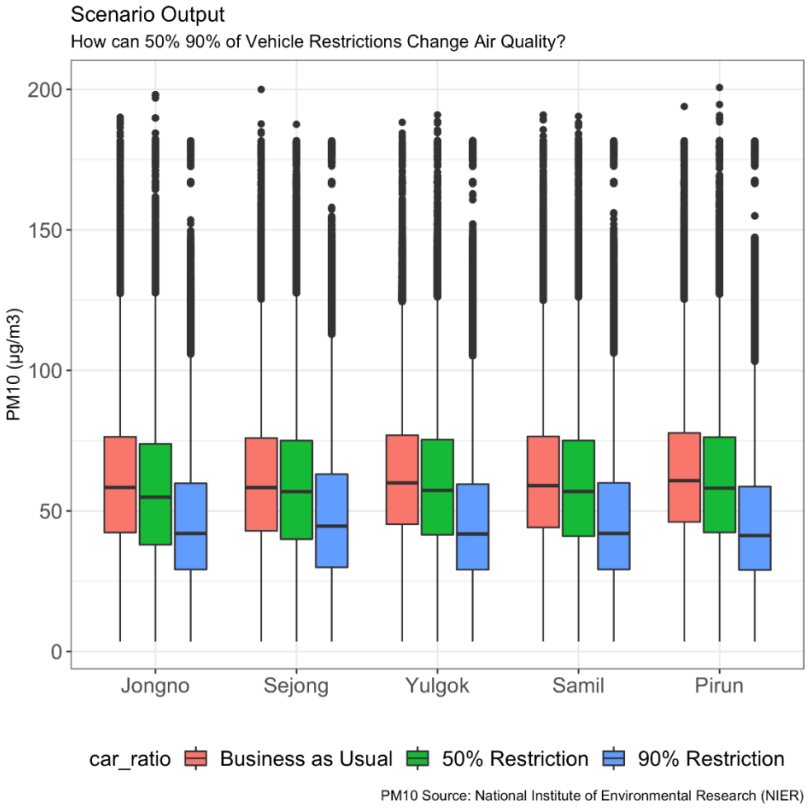


Fig. 5.30 Boxplot of PM_{10} by each restriction scenario

drivers. When the subway commuters walked the shortest route (awareness "no" scenario), the group showed several occurrences of onsets and peaks where the highest values were around 8% (see Figure 5.31). Fewer occurrences and lower peak values (5%) were observed when people paid more attention to avoid high PM_{10} values (awareness "yes" scenario). There was a trade-off between shortening travel distance but allowing oneself to be exposed to high PM_{10} and minimising the exposure levels but end up spending more time outside. This chapter tested the sensitivity of walking speed, taking speed 1 down to .2, but discovered only a 1% increase to the peak.

The awareness scenario was also not as effective for resident drivers. While it was assumed that an unwell population (nominal health below 100) stayed at home for recovery during the weekend, not on weekdays, the resident drivers experienced a marginally lower health risk when the scenario was activated (comparing the first peak of BAU for each scenario: 9%→6%). The risk rate also decreased when more stringent vehicle restrictions were applied to the scenario (9%→5.5%), however, not

much difference was indicated between the awareness scenarios under high restriction scenarios.

In comparison with subway commuters, fewer onsets were discovered for the resident drivers; however, on March 25th when the PM_{10} exceeded 100µg/m^3 for two full days, the risk population skyrocketed to 86.7%.

Given the conditions, the outcome unfolds that although a 20% decrease in PM_{10} was identified through tighter vehicle restrictions, changes in pedestrian behaviour did not seem to allow for a significant mitigation of exposure in both groups.

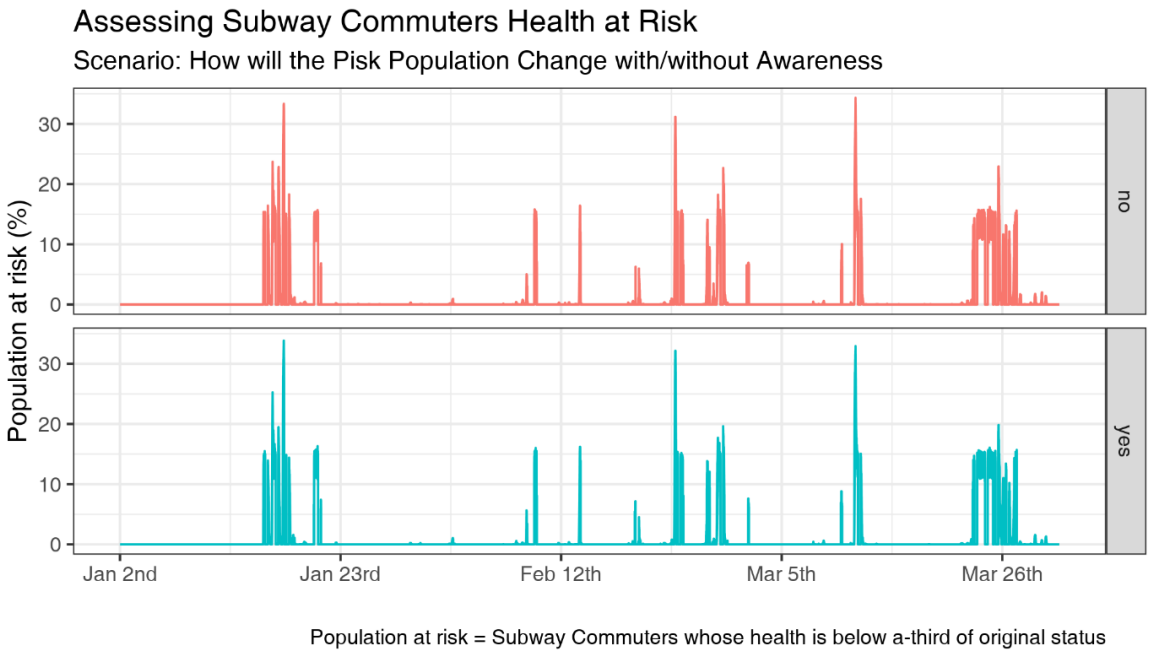


Fig. 5.31 Health risk of subway commuters compared in the awareness scenario

5.7 Summary and Discussion

This chapter constructed a traffic simulation for central Seoul to investigate the coupled problem of NEE and exposure to PM_{10} in groups of pedestrians and resident drivers. Overall, significant extra particulates were found to exist along roadways. Although longer exposure times for pedestrians led to a larger accumulated exposure overall, the majority of drivers were exposed to the highest levels of pollution (>150µg/m^3), which

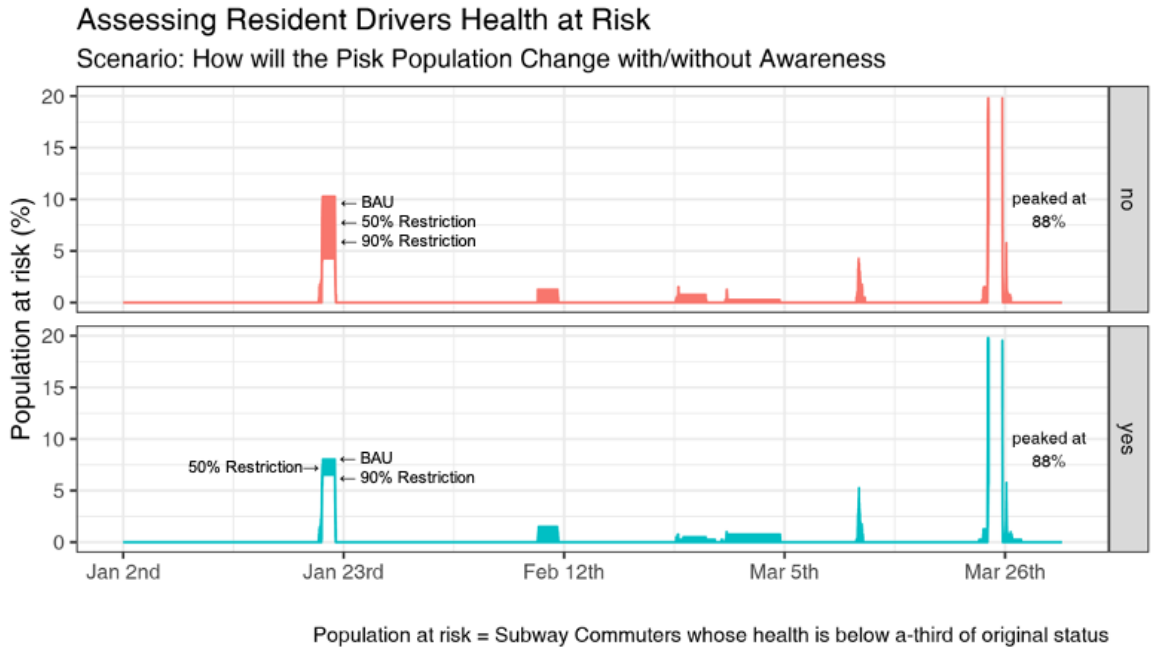


Fig. 5.32 Health risk of resident drivers compared in the awareness scenario

was largely due to the time spent in congested areas. The health effects, however, depended strongly on how the impact and recovery from exposure were parameterised.

Differences in health effects between pedestrians and drivers. The chief difference between the two groups is the exposure-response to a sudden PM_{10} rise. Even on a highly polluted day, on January 20th for example, only 7% of resident drivers experienced an acute health risk, while nearly every subway commuter was unwell but soon recovered as they arrived at their workplaces. However, when high PM_{10} concentrations lasted for a couple of days, such as March 24-25th, the result was the exact opposite: 15-20% of pedestrians had health risks related to high PM_{10} exposure, whereas 88% of drivers became unwell. This evidence suggests that exposure time to ambient air pollution may affect a severe health risk to some vulnerable people, but the duration of extreme air pollution with vehicle emissions can end up having significant risks to health (Laumbach et al., 2015).

Travel time was also an important element that differentiated the exposure patterns. Subway commuters were assumed to travel at a certain time after 6am. The fluctuations in the risk population for subway commuters occurred due to the variation and duration of the commute period. This can extend the study of Gurram et al. (2019) that

associated long commute time with higher exposure: the commute of rural passengers with long commute time (>60 minutes) had 8% higher exposure of NO_x than those whose daily travel time was <30 min.

A reduced risk for drivers occurred because of the model setting that drivers only move from indoors to indoors (e.g. home-car to cabin-office) had a different air filtration ratio. Although extreme PM_{10} was identified, cars can act as a semi-sealed environment to prevent toxic substances from getting in, which can sometimes be safer than the environment for pedestrians (Briggs et al., 2008; Gulliver and Briggs, 2007). Gulliver and Briggs (2007) found from 33 pedestrians and 33 in-car drivers from Leicester, UK that the average walking to in-car ratios were 1.5 for intermediate (PM_{10}-$PM_{2.5}$) during January-March 2005. Briggs et al. (2008) collected particulates whilst walking and driving 48 routes in London, and discovered that the average walking to in-car ratios during May and June 2005 were 4.7 times higher for intermediate (PM_{10}-$PM_{2.5}$). The study revealed that staying inside the vehicle with the windows closed will filter out the immediate ingress of particulates, and more importantly the journey time between walking (7-30mins) and vehicles (1-15mins) was the most critical factor linked to higher particulate exposure. Findings from Odekanle et al. (2017) and Żak et al. (2017) oppose that staying inside vehicles face higher exposure and health risks than that of pedestrians. Odekanle et al. (2017) measured carbon monoxide and VOC in Lagos, Nigeria for 12 days during the morning and afternoon rush hours, while Żak et al. (2017) collected a day-long data from different sample points of Poland cities where vehicles and pedestrians have passed by. However, this chapter criticises that the study periods were too short to generalise the results, and a passive way of pollution sampling does not clearly distinguishes the levels of particulate exposure between driving and walking.

The conflicted findings lead to a conclusion that exposure levels strongly depend on the type of pollution and study design. Thus, to strengthen this argument, there must be a supporting dataset that traces the same route between drivers and pedestrians and measures pollution.

The contribution of NEE to ambient PM_{10}. Road traffic was found to have contributed around 35% of the average roadside PM_{10} concentrations, with much higher contributions on a finer time scale. The rise of PM_{10} was largely due to the amount of traffic that emitted NEE on roads, regardless of the fuel type and mode of power (Air Quality Expert Group, 2019). These findings are consistent with the case

study of Weinbruch et al. (2014), where 48% of roadside PM_{10} in Ruhr was contributed by particulate emissions, but because the study merged exhaust and non-exhaust emissions for traffic sources, the rate of roadside PM_{10} might be closer to the findings of this study.

The selection of parameters and their sensitivity. Since the behaviour of vehicles and pedestrians is important for agent-based simulation, this chapter selected parameters that generated emissions and those that related to individual health. First, emissions from vehicles consist of tyre wear, brake wear, and road surfaces that contain sub-parameters, which for tyre wear, for instance, include many vehicles within the defined spatial boundary, mass fraction of particles (g/km) and travelling speed which all have different units (see equations introduced in Section 5.4.7). Among these, the number of cars was the appropriate indicator for use because the other indicators were affected by the conversion of spatial units from kilometres to 30 metres.

Dilution, or the duration of NEE before it mixes with the local atmosphere, was very sensitive to the parameters. As the dilution parameter increased, the overall PM_{10} increased significantly by a maximum of 11µg/m^3 on `Emission 20`. PM_{10} remained higher on longer duration because the patch that carries the emissions only disappears when no vehicles approach for $10+\alpha$ ticks, which is impossible before the night-time hours. In other words, most roads will experience high pollution apart from between 2am and 5am. However, the range of dispersion did not seem to affect it as much as dilution did. This is partly because the angle of the pollution cone only differed by one patch, between 45° and 90°, which did not seem to be sensitive. However, as mentioned, the wider dispersion can inflate the PM_{10} concentration to nearby pavements that will be exposed to polluted air. The sensitivity was also small with respect to the ratio of non-resident cars where only a difference of <1µg/m^3 was seen between 2.5% and 10%, while 10µg/m^3 less was observed when no extra vehicles were introduced into the study domain. Contrary to the study's expectations, the overall PM_{10} decreased as the ratio exceeded 10%. Although queues on the Samil road (lower-central entrance) and the Yulgok road (northeastern entrance) were seen in the 2.5%, 5%, and 10% ratio simulation during the daytime, the queues suddenly extended greatly in the 20% ratio simulation, with the result that the PM_{10} became lower than for the smaller samples. This happened due to the capacity limit of the software that assumed vehicles maintained a one-patch distance from the vehicle in front and allowed two lanes, one for traffic in each direction, separated by the heading function.

With regard to health-related indicators, health loss was highly sensitive for the at-risk population. As mentioned earlier, several fluctuations in risk rate were seen among subway commuters during major pollution episodes, but the maximum at-risk population showed an exponential increase between α (parameters) 0.1 and `0.15` that resulted in 30% and 100% of the at-risk population respectively. α was applied as an exponential component in health loss; thus, a small difference between values can change the outcome excessively, despite the support of health recovery. The result of zero cars did not show a large difference. This was mostly because pedestrians walked through quieter areas than near roadways, which enabled pedestrians to avoid highly polluted areas (European Lung Foundation, 2020). Drivers also experienced a large difference in risk population over time, where 5% suddenly rose to 50% between α 0.1–0.15; however, a remarkable decrease was also seen when non-resident cars were absent (15%). Although studies found no evidence of high risk related to particle exposure inside cars, many studies that examine indoor-outdoor ratios for pollution have proven that opening a window can contaminate indoor air pollution (Kreider et al., 2019), which can also be a useful guideline for vehicles.

Slower walking speed resulted in a significant rise in the risk rate, but fast walking only made a small difference. Compared to the default walking speed of 0.6-1.0 patch/min., the result of slow walking (0.2-0.4 patch/min) resulted in a 10% higher risk population in extreme pollution episodes. By contrast, over 1.5 patch/min made little difference from the default speed, which means that even rapid walking cannot help in avoiding instantaneous exposure to a sudden rise of PM_{10}. The result is very meaningful because the arguments that "faster walking and breathing at a higher rate over a shorter duration of exposure can significantly reduce the absorption of pollution" might not be taken for granted in this study in which exposure to a sudden pollution rise is critical to pedestrians.

How effective were the scenarios?. In the vehicle restriction scenario, roadside PM_{10} showed a 20% decrease when the majority of cars (90%) were banned from the city centre, consistent with other estimates of the contribution of NEEs. It was already striking to see an overall decline in all the roads by $10\mu g/m^3$, but this was because less extreme values, mainly from traffic, were observed. Similar results were introduced in Munich where PM_{10} within the great ring road was reduced by 2-10% after the Low Emission Zone was enacted (Cyrys et al., 2014), while London experienced a great

fall in NO_2 but not in particulates PM_{10} or $PM_{2.5}$ (Mudway et al., 2019; Wood et al., 2015).

Changes in pedestrian behaviour did not seem to allow for significant mitigation of exposure (e.g. a smartphone app indicating areas of high PM_{10}). Pedestrians who travelled on a highly polluted day had significant health damage, and because the concentration in the background areas was comparable, it was less likely to make a remarkable difference to the outcome, unless the person stayed at home. While some may assert that exercise outweighs the harm from exposure to pollution rather than staying at home (Tainio et al., 2016), another study opposed the idea and stated that avoiding pollution exposure is important for susceptible individuals who have chronic cardiovascular or pulmonary diseases and should be recommended when pollution levels are particularly high (Laumbach et al., 2015). Although various findings lead to unending debates, careful consideration needs to be made of the design of the study (long-term or short-term, average exposure over a certain period or instantaneous exposure) and the scope of the participants.

Strengths and limitations of this study. This study was the first to jointly examine the contribution of NEEs and the adverse health effects on a group of commuters from a microscopic approach. The application of NEEs on a patch level sufficiently recreated the generation of brake and tyre wear and the following dispersion that can possibly happen in the real world. Although the equation was cited from studies that used distance-driven indicators (Breuer et al., 2020; EMEP/EEA, 2019), this study simulates a real-time method of emissions and dispersion from each vehicle to better articulate the causation between the polluters (vehicle emissions) and the susceptible people (drivers and pedestrians).

However, this study also has several limitations. First, this simulation has constraints in its spatial boundary of Central Seoul and the temporal boundary between January and March. Although the study area is known for heavy traffic and the largest population differences between day and night, it only represents 2.5% of Seoul. In addition, winter in Seoul is known to have the highest PM_{10} episodes which occur frequently. Hence, this simulation is calibrated to the winter season where the temperature is below 0°C with less humidity and more heating, and vehicle emissions are thus expected.

Second, the study used a collective sample of inbound vehicles to mimic the flow of traffic within the spatial domain, but the coarseness of traffic signal settings and the

interaction between vehicles in front of the signal failed to control the traffic effectively. It normally takes 1-2 minutes for a vehicle to pass a single road segment, but since the model did not consider multiple lanes in the segment but was given a stop in front of traffic signals and a slow-down function to vehicles ahead of their directions, more vehicles queued up outside the spatial boundary.

Third, NEE was dispersed to the neighbouring patches and diluted after a duration of minutes. However, the current model is assumed to have very stable and still airflow with no wind, which is realistically untrue. In fluid dynamics modelling and atmospheric modelling, such as the ADMS model (Beevers et al., 2013), wind direction and speed are considered to be important indicators in pollution dispersion because wind can determine the patterns and trends of local air movement.

Fourth, although health loss was parameterised, there is no suggestion that the parameter values are warrantable unless they have been calibrated and referenced by known parameters. The current model tested five health loss parameters from `0.03` to `0.2` and observed that the acute health risks were always 100% in every pollution episode when the parameters were over `0.15`. This seemed unrealistic. Even though the model is illustrative, the absence of accurate health loss parameters might not fully articulate the acute health risks from short-term exposure.

A final limitation of this study was in generating the background PM_{10} levels from two urban background stations in the study area. Despite a large pollution variation expected in urban areas, particularly where heavy traffic and high-rise buildings trap the airflow, this study only applied non-road patches to the simulation from two background monitoring stations. This may have overlooked the geographic variation that is commonly noticed by personal monitors (Norton, 2015).

Future works. Perceiving the severity of NEE to ambient PM_{10}, future studies should consider more sophisticated ways that particles from brakes and tyre wear are generated and dispersed on roads, and whether the shortest-distance trip on major roads is better than taking a less congested route but one with extended travel time in terms of reducing one's exposure level. In addition, immediate improvements can be made with respect to domain expansion to prevent vehicles idling outside the study domain, which will also provide more opportunity for geographic variation of pollution attributed to traffic volume.

From a health perspective, it may also be useful to model other working groups in the CBD area. Taxi drivers, for example, can be a good occupation to study, as many drivers in Seoul work for up to 12 hours inside a car, but little is known about their exposure level, history of the disease, age, or working patterns, which vary immensely. This chapter can also blend the contents with Chapter 4 to specify the movements between transport modes and demographic subgroups. The incorporated model will indeed be more complex than the individual models but the delicacy of health loss functions by demographic groups together with the in-vehicle operations can suggest how different demographic groups can be more/less exposed by their commute patterns.

Chapter 6

General Discussion

6.1 Introduction

Health outcomes related to urban air pollution exposure is an ongoing problem which needs to be investigated both at a micro and macro scale (Caplin et al., 2019; Reames and Bravo, 2019). Until recently, tracking individuals for a continuous amount of time was almost impossible, and even where it is now possible, one may run into problems with devices needing continuous power. Even if individual tracking can be monitored, it may take around 30 years to get a meaningful causation between exposure and health outcomes over a large segment of the population, with diverse individual characteristics. Without individual measuring devices, the interpolated pollution fields with sparse fixed-site monitors cannot be meaningfully interpolated down to the individual scale. Another problem of large and smoothed interpolation is that it might ignore the spatial and temporal peaks of pollution that might be vital to health outcomes. Consequently, there remains a mismatch between the large scale of interpolated mapping analysis and the individuals being exposed to air pollution from their activity patterns.

To estimate population exposure to particulates from a bottom-up approach, this thesis develops a conceptual framework to contextually understand air pollution and health vulnerability through the variables of exposure, sensitivity, behaviour, and adaptation (Jerrett et al., 2010; Parry, 2007). The data and findings arising from Chapter 3 revealed that although both spatial interpolation methods created a comparable pollution outcome over Seoul, the chapter was able to illustrate the spatial variability of extremes in different parts of the city (i.e. which area was

continuously higher/lower during the study period), and more importantly how the spatial heterogeneity introduced by the road system can apparently manifest roadside pollution within a short amount of execution time. The thesis applies ABM to improve on some aspects of the spatial and temporal dynamics from previous studies. Chapter 4 used a sample population of 25 districts and simulated home-work movement for 12 years to evaluate the health risk between demographic groups. The findings reveal that disparities in health may arise as a result of differences in demographic groups; wealthier districts or districts that have fewer inhabitants near roads have less risk under future scenarios. Chapter 5 applied a finer scale movement on drivers and pedestrians to simulate the dynamics of activity patterns and use vehicles to generate non-exhaust emissions to the local atmosphere. The chapter found that around 40% of the roadside PM_{10} can be contributed by non-exhaust emissions, but the exposure levels depend on the trajectories between individuals, but was substantially affected by the days when extreme pollution was observed.

Since the empirical chapters contain separate discussions of their analyses, this chapter addresses discussions that will take this thesis towards a broader perspective. The first section comments on the data biases and absences that were discovered during the data collection process. The following section points out the unsolved problem of the exposure-response relationship that is difficult to observe without longitudinal tracking of patients. The third section highlights how microscopic simulations can be useful for future health applications. The chapter concludes by writing about the benefits and future directions of open source research.

6.2 Data Biases and Missingness

Exploring biases and missingness was one of the key tasks of the thesis. This section highlights how the thesis disentangled missing data and potential data biases for 1) pollution and 2) population activity patterns.

Pollution Data

Earlier in this study, two questions arose from analysing the original data. First, how will this study handle missing values in time series across the stations? One way

to harness missing values particularly for time-series was either using a statistical imputation. Imputation does not necessarily give better results because the value can be abnormally high compared to the past and future observations, or the missing gaps may be so wide that values depend heavily on the hypothesis, e.g. PM_{10} values in February were between 40-1500µg/m^3. Since the pollution data for Chapter 4 used pollution data from the nearest district, and each patch chose one of the values from working hours at every time step, using the imputation method allows ways to create more options to pick pollution values at random for each spatial location.

This study tested various methods including Last Observation Carried Forward (LOCF), Moving Average, (linear) Interpolation, and Kalman Filter Sampling (KFS) via the R package `imputeTS` (Moritz, 2019). Amongst these methods, KFS seemed to be the appropriate method because the Kalman filters find the records for the same (seasonal) period in previous years with autoregressive integrated moving average models (ARIMA), then creates the best estimate to fit the trend (Hadeed et al., 2020). However, there was no rule-of-thumb in terms of output accuracy especially when the missing values appeared for the long-term. A month-long gap was observed at Gangnam monitoring station during February 2012, presumably due to regular inspection and relocating the station near the district hall (see Figure 6.1). Having compared the values between imputation methods as well as the nearest districts, the differences were not worryingly different (<1% for the entire period). While there was no consensus in the literature regarding an acceptable rate of missing data in a database, Schafer (1999) mentioned that an "inconsequential" rate would be 5% or less.

The second question was how can data biases occur from spatially sparse monitors? The study noticed that spatial interpolations would usually underestimate the outputs at an unknown region. This is because the predictions are normally judged by the mean and standard errors, and thus do not fully articulate the extremes that can happen on roadsides, for example in Baxter et al. (2013), Dias and Tchepel (2018), and Munir et al. (2019). To overcome the over- and underestimations, Chapter 3 compared Universal Kriging and Generalised Additive Models with a sample of 57 stations across Seoul (605km^2) and weighted additional pollution on road networks based on the observations retrieved from roadside monitors. However, interpolating a wide area with only 57 data points gave a rather smoothed structure for a city scale, which does not consider small-scale spatial variance, i.e. fields have similar values over small areas. Statistical models have an advantage for quick execution over a larger

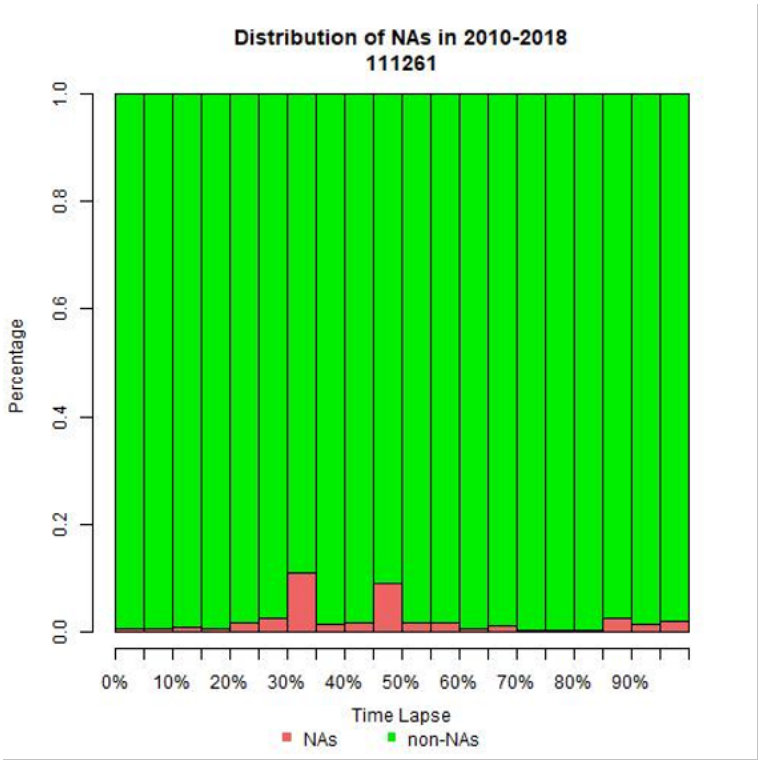


Fig. 6.1 Missing value rates of PM_{10} at Gangnam Background Station (2010-2018)

space; however, a simple weight on the road network will not sufficiently account for vehicle emissions and meteorological factors that are important in modelling street scale pollution (Dewulf et al., 2016; Smith et al., 2016).

Chapter 5 also used the background data to generate PM_{10} patches across the study domain but made a step forward from previous models by simulating non-exhaust emissions and diffusion to the roadside. Having calibrated the model with roadside PM_{10} observations, PM_{10} on roadsides seemed to fit well with the observations but the study needs to be aware that the concentrations were very sensitive to emission and dilution parameters. Other untested parameters that include the number of lanes and traffic signals may well change the outputs to a large extent.

For future works, this study can consider crowdsourced data to improve both data biases and absences. Low-cost sensors (Castell et al., 2017), Twitter or web platforms (Zheng et al., 2018) can afford a unique opportunity to provide information over a shorter time as well as over a wider range of spatial patterns (Schneider et al., 2017; Zheng et al., 2018). In practice, as one of the citizen science projects, Seoul has initiated

a 'SUM Seoul' project – SUM as an acronym of Smart Urban air quality Management and means "breathe" in Korean – to install over 500 smart sensors to measure real-time $PM_{2.5}$ and PM_{10} pollutants at schools and in deprived neighbourhoods. Denser data points across the city would not only allow researchers to calibrate their models to generate more realistic results to improve exposure science, but also means that collaborations with stakeholders as well as citizen participation can help tackle air pollution in our neighbourhoods.

Activity Patterns

Absences in data were also discovered in population movements for Chapters 4 and 5. Chapter 4 assigned each demographic subgroup different activity patterns, where the young and elderly were assigned to nearer patches and the economically active moved to their destinations based on OD matrices. Since the OD matrices were based on the annual average daily traffic 2015 (AADT), the model had to assign a home-work movement to move the whole population but exclude the trip routes. Although this was a good place to start, neglecting trip routes cannot accurately account for the time spent outdoors or indoors nor compare exposure levels by transport modes (for example, (Shin and Bithell, 2019; Smith et al., 2016)).

To overcome the problems of using OD matrices in Chapter 4, Chapter 5 simulated agents' trip routes using a minute-by-minute time scale. This study found that adding trip routes on a finer time scale has aided a more intuitive understanding from what time and where the person is commuting to, especially on a highly polluted day. Also, applying activity patterns was effective to distinguish exposure patterns between transport modes, where health risk was more frequent on subway commuters due to the time spent outdoors, while resident drivers were less frequently exposed to extreme pollution as they spent most time moving between indoors.

Chapter 5 assigned the shortest route on road links for vehicles prior to the implementation using the A* algorithm. However, assigning the shortest route for all vehicles might entail another problem of traffic congestion because they only have one route in mind. In reality, when drivers encounter traffic congestion they may either join the long queue with other vehicles until the congestion fades out or choose one of the alternative routes to reach their destinations but both decisions would depend on

the trade-off between time and distance. Thus, to overcome the limitations of using the shortest path algorithm, future improvements can consider cognitive behaviour modelling. One of the approaches is called "fast and frugal". Simply speaking, fast and frugal is a decision tree that aims to find a goal from many queries[1] and rapid execution. Rather than a single solution, the decision-tree algorithm will allow a 'bounded rationality' for agents to decide their routes (Hafenbrädl et al., 2016). That is, allowing space for the individual itself to choose directions based on its memory of spatial knowledge about the surrounding environments. Manley et al. (2015a) have applied a fast and frugal tree to suggest various road routing options with three spatial hierarchies: region, node, and road scale[2] (see Figure 6.2). They mentioned that the difference between the methods is the way of calculation (i.e. whether it starts to eliminate unnecessary areas first, or just seeks alternative routes based on the nearby node/roads). Also, in the pollution context, this study can consider more parameters in the decision tree to suggest myriad options for vehicles to perceive: possible congestion points, high polluted areas, or considering road hierarchies. For the spatial scale of simulation used in this study, obtaining more options of navigation can be useful for finding the shortest distance, but more broadly, this approach can provide insights on how human cognition and spatial knowledge can improve the heuristics for pathfinding solutions.

6.3 The Unsolved Problem of the Exposure-Response Relationship

Chapters 4 and 5 have applied ABM to estimate how exposure to PM_{10} can cause severe health outcomes. In toxicity science, the exposure-response relationship is explained as "an increase of exposure associated with an increased or decreased outcome", which means that there may be a causal relationship between exposure and response (Pettygrove, 2016). However, the lack of information to prove 'response' as well as using ground-monitoring sites to measure 'exposures' can lead to great uncertainty.

[1] In psychology, queries are termed as cues, e.g. how many junctions does the car meet between the origin and destination?

[2] A region-based choice removes all the unnecessary possibilities and finds the best link according to the highest score. Node based approach preselects a series of nodes that connect between origin and destination, but here the agent moves to the next node considering the minimum angular deviation between the current node and the goal node. Finally, a road-based approach chooses the routes based on the minimal distance measure and presumably the hierarchy of the nodes.

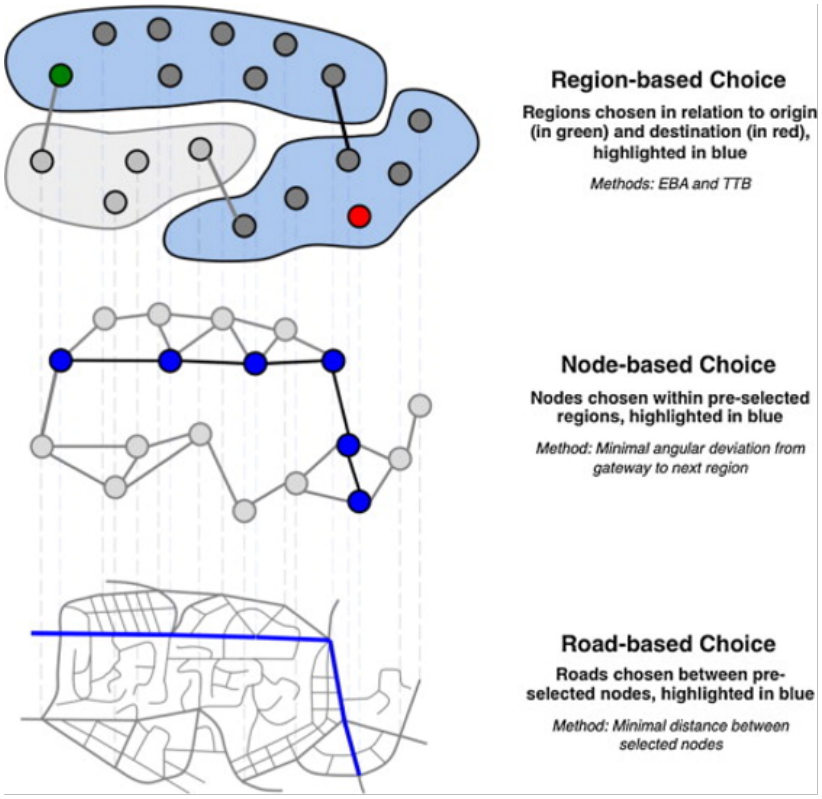


Fig. 6.2 A conceptual illustration of the heuristic route choices across different hierarchies (Manley et al., 2015b). [Top] Region-based choice assigns coarse route plans by excluding potential gateways in a step-by-step process: excluding all potential gateways that do not match the criteria (EBA, eliminate-by-aspect), then by calculating total score according to the highest score (TTB, take-the-best-region). [Middle] Node-based choice is the simplest and less cognitively intensive way to connect the origin and destination. [Botttom] Road-based choice finds routes based on the minimal distance measure.

One of the obvious reasons is that most people stay indoors (Smith et al., 2016). While this thesis attempted to simulate the process of exposure-response with ABM, the epidemiological causality is still an open question.

This thesis examined two cases of exposure: long-term exposure (Chapter 4) and short-term exposure (Chapter 5). Both studies have applied the same threshold model to degrade health whenever the PM_{10} on the patch the individual was standing in exceeded $100\mu g/m^3$. Once the individual's response reaches 1/3 of the initial status, the individual's health changes from a normal to an at-risk status. Although the ratio

1/3 is arbitrary and the threshold was only illustrative at present, it aimed to emulate the latent period from exposure to a health outcome, which was a good place to start.

However, this thesis has potential limitations in relation to health outcomes. First, this study discovered that the initial conditions strongly affected the outcomes. Even though all agents had heterogeneous characteristics and behaviours and moved to different places, the scenario outcomes appeared almost simultaneously within the subgroups. This is purely because the agents were designed to have no previous exposure to air pollution, i.e. all agents started with nominal health of 300. If health history was provided at least for a sample of the population, it would have been more appropriate to suggest health promotion practices with better evidence. To the author's knowledge, there was only one case study that showed the direct relationship between NO_2 exposure and an asthma attack (Hancock, 2020). This was the death of a 9-year-old British girl after suffering from asthma for two years, and there was a record that NO2 in front of her house exceeded 40µg/m^3 whenever she visited the hospital due to breathing problems. No significant association would have been revealed if the patient did not visit the hospital for regular check-ups.

Future work In this regard, future studies can consider using longitudinal data to associate long-suffering patients who are symptomatic to air pollution (Van Roosbroeck et al., 2007). The immediate scenario is to obtain multiple years of data (preferably consecutive years) to adjust the health loss and health resilience parameters, but since the cost of obtaining data to conduct research is around £200, along with the agreement to comply with the confidentiality policy, this also depends on the funding available. Alternatively, if this research can collaborate with Biobanks, such as the UK biobank for example, then it can surely deal with the procedure of health disorders by tracking family history, lack of nutrition, or type of occupation (Checkoway et al., 2007; Özkaynak et al., 2013).

For interdisciplinary work, researchers can collaborate with epidemiologists to understand more about the causal relationship between exposure and response. Özkaynak et al. (2013) stated that to examine the causality, the researcher needs to seek a disease attributed to air pollution exposure. This is because the type of disease will enable the selection between acute and chronic diseases, and once the disease is determined a short-term or long-term exposure model can be established. This reversed approach has been recently facilitated with the concept named Inverse Generative Social Science (IGSS). The motto of IGSS is, "if you didn't grow it, you didn't explain it". It means

the modeller needs to grow/suggest as many ways as possible to explain a societal phenomenon (e.g. crime, disease) (Vu et al., 2019). Unlike the traditional ABM approach that follows the process of design agents → assign rules → to verify and calibrate the model → return outcome, the IGSS follows the motto by going back to the 'explanation (outcome)' and finding other ways of growing it. The questions can be enumerated as follows: 1) what are the main health disorders as a result of pollution exposure?, 2) if the disease was chronic, how can we use the medical history of patients to complete the set of rules?, and 3) how robust were the outcomes to a small change of parameter adjustments?. These questions will offer a variety of answers to this model and lead a practical way from what Özkaynak et al. (2013) were aiming for.

6.4 Prospective Avenues for Future Health Applications

ABM scenarios can project outcomes from global (e.g. air pollution trends) to local parameters (e.g. individual behaviour). ABM scenarios can identify leverage points where small adjustments can generate large shifts in systematic outcomes, which are known as 'tipping points' (Hammond, 2015). These all are advantages which are difficult to conduct in *in vitro* experiments (Hammond, 2015; Wilensky and Rand, 2015).

Based on these features, this study used scenarios with a futuristic aspect to understand:

- 'What if the pollution trend increases/decreases/is maintained?' (Chapter 4)
- 'What if the agents' health resilience is a small/large effect?' (Chapter 4)
- 'How will a vehicle ban from the city centre improve air quality and health risks?' (Chapter 5)
- 'How would the trajectories of lower air pollution help mitigate health risks?' (Chapter 5)

The first three scenarios focused more on parameter adjustments, while the last scenario controlled the behavioural features of individuals. Despite interesting results,

the weakness of this study was that most of the exposure was determined by the exogenous parameters, but did not use the microlevel configurations of agents that ABM is good at.

Looking at other applications, there was another threshold model for ABM that investigated how school students change their protective behaviours in response to an influenza epidemic (Badham et al., 2018). Students changed their behaviours when the sum of personal attitude, the perception of norms, behaviours of others, and the probability of new infection exceeded the threshold. Here, the norm was the nearby individuals who have adopted protective behaviours such as washing hands frequently. The model allowed the agents to interact and share information with friends, teachers, and social media, but at the same time, allowing physical contact has induced the possibility of contracting influenza. If somebody had contracted a new virus, they will abruptly change behaviours to protect themselves. This study did not simulate the interactions of students in an intuitive manner but rather used ABM to give some uncertainty in parameters to produce the outcome. However, it was useful to understand how psychological variables can be quantified and reflected to simulate behavioural adjustments.

For future works, this thesis can take into account more activity options such as 'work from home' for employees, and 'indoor activity only' for students at school when pollution is above the threshold. Additionally, if possible, this study can add a communication parameter to share information before the activity. For example, if PM_{10} is considerably high for a week (e.g. a daily PM_{10} average $>200\mu g/m^3$), schools and firms can decide whether to let people come to school or work; but since pollution varies by district, the strategy of avoiding air pollution can be different between schools, firms, and districts.

Regarding health loss and resilience, health resilience scenarios have not fully considered the variation in recovery speed by demographic groups. Due to the lack of personal health recovery, it was uneasy to give heterogeneous resilience apart from the age effect. In general, old people have a comparatively slower recovery speed than young people as their physical metabolism has aged over time. However, there might be exceptional cases within the group who have sustained their health or vice versa. A possible way to seek for future studies is to obtain patient data and analyse the length of admission. It might not be perfect data, but this will at least hint at the likelihood of health recovery due to pollution exposure.

6.5 Technicalities: The Benefits of Open Source Research

This study took a combination of data-driven and simulation modelling approaches to investigate this thesis. As such, there were technical processes from downloading the data to produce outcomes, which are not normally shared in dissertations and journal articles. However, with the advent of the era of big data and data science where people collaborate online and the atmosphere of 'sharing is caring' has become a norm, opening up research would no longer be an option. This section discusses the benefits of conducting an open source project and suggests some things that researchers should keep in mind.

Benefits of Aiming an Open Source Project It is argued here that being transparent and honest about your model development strengthens the credibility of the research. In doing so, this thesis shared the original data, the scripts with annotations that are downloadable and executable, and wiki pages to summarise the outcomes and interpretations (see Figure 6.3 for example). This evidence enables scholars and technicians to visit the repository if they are interested in the source codes or outcomes. Also, people can comment if any errors or bugs are identified or the model is not executing on their machine or may suggest alternative ways to tackle the same problem. Even during the development, many developers share their work via online repositories (e.g. Github, Gitlab), and Zenodo to ask for advice. Agent-based models are mostly uploaded on OpenABM (comses.net). All of this can improve the quality of research.

More practically, one can learn new ideas by helping each other. If there was a technical issue which can't be solved, the problem should not be kept hidden, but rather be opened and solved together with experts online and offline. Figure 6.4 is a pragmatic example from this during the agent-based modelling process. The experiment was to produce a spatial outcome on every tick in `R` (Salecker et al., 2019). Since R stores all the assigned objects in the RAM, producing 96,822 patches over 8764 ticks made even the HPC (High-Performance Computing) clusters struggle. Developers from an online community named *Stackoverflow* also informed that NetLogo itself has memory limits, around 1GB, and the software keeps each run in the memory before it shuts down, thus running models in a parameter-space exploration (e.g. behaviour space) which can decrease the execution speed. Before this information was seen, it was not understood why the model took 1 hour 20 minutes to finish the first run but struggled to maintain

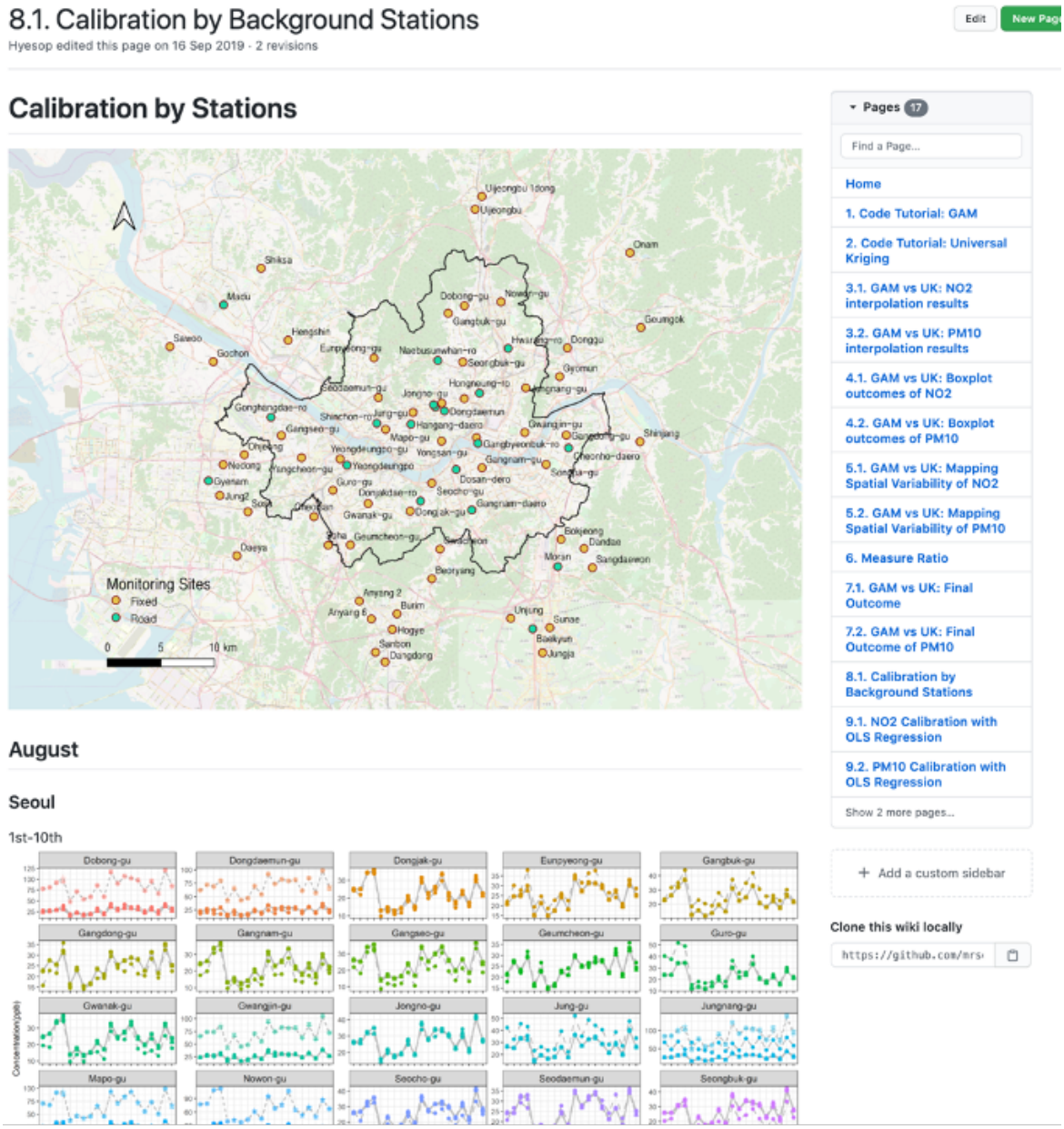


Fig. 6.3 A screenshot of a Github page showing how open platforms can help other people to understand the outcomes step by step

that speed after the thirtieth run. Hence, sharing technical obstacles that occur in the middle of research can save a lot of time even for those who are contemplating similar research.

The Future for Open Source Research This study suggests that for future quantitative studies in social science, students and researchers in academia should

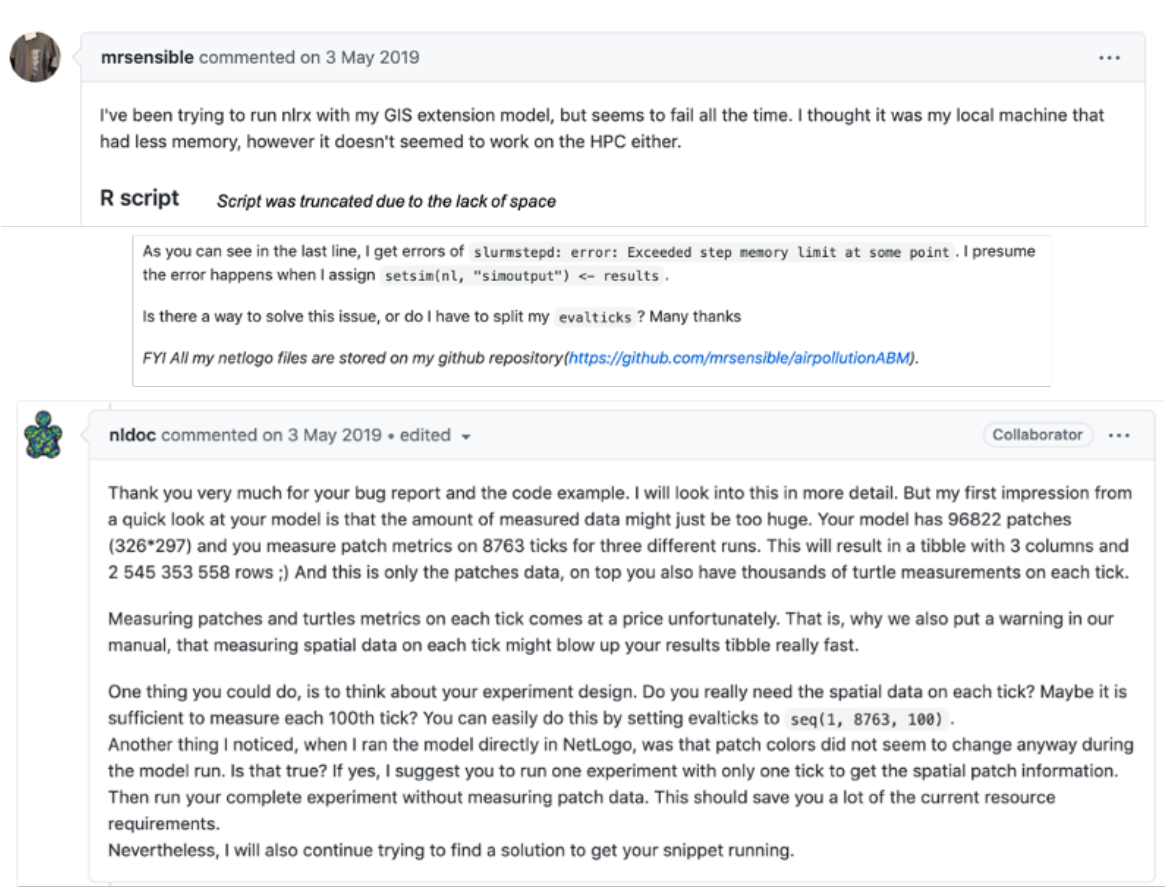


Fig. 6.4 Comments posted on an online repository regarding the memory issue that NetLogo and R encountered

acclimatise themselves to using open source platforms to conduct sustainable research. Similarly to good writing skills such as clarity, conciseness, and coherence, good programming should take into consideration the following points. First is clarity and conciseness (CC). Here, clarity means that the scripts should be neatly documented. The computer does not know whether the codes are dirty or neat, it only cares whether it is syntactically correct, but it matters when other people attempt to understand the task. If the outcome produces the same results, it is always better to write clearer and simpler codes for other people and future upgrades. Thus, researchers should refer to other people's work and learn how to code effectively. Another way to maintain clarity in coding is to keep descriptive and distinctive names for new variables. This statement might seem contradictory to the conciseness issue, but this is important as one of the common mistakes users make is to assign variables with abstract names such as `LP1, LP2,...,LP10`, which seems clear and concise for the model builder, but is even harder

for the others when reviewing the code. The famous quote, "Everything should be made as simple as possible, but not simpler." Einstein (2010) is the appropriate phrase that model builders should always keep in mind. Hence, instead of coding `LP9`, names such as `LandPriceIncreaseRate2009` (camel cases) or `landprice_incrate_2009` (snake cases) can be more effective for the reviewers to understand the model.

Second is reproducibility and replicability (RR). To be reproducible and replicable, initially, no errors should occur when others execute the script, and possible errors or bugs should be reported. It will also be more useful to document the libraries and the dependencies required. This is quite important as different OSs (operating systems) have different behaviours to install packages. For instance, the `sf` package in R has slightly different ways to install the package between OSs where Windows and MacOSX can be installed from the binary package while Linux needs to separately install GDAL (to read and write vector and raster data), Proj (which deals with projection), and GEOS (which provides geospatial functions) prior to the package installation. Finally, it would be very helpful if unit testing is included in the model. While R and Python provide splendid examples in their vignettes, NetLogo remains to offer the library models but goes no further than that. Offering unit testing examples can give a better understanding when the whole model is too complicated for others to comprehend. It can also give the impression that the modeller has full control of the model because without the unit test the verification process becomes error prone.

Third is to maintain version control. In terms of sustainability, researchers should be aware of software maintenance. Much programming software relies on libraries and packages that are built on a particular version. If the software is upgraded and no longer accepts the previous versions, then the package developers need to keep updating to run it on a new version. For example, *NetLogo 6.0* experienced a significant change compared to versions 5.X. The biggest change was the replacement of tasks by anonymous procedures(Wilensky, 1999). This means that tasks are no longer primitives but are converted to arrow syntax. For example, if there is a list of `[a b c]`, the previous task is asked to add the first, second, and third element as `foreach [a b c] [ ?a+?b+?c ]`, while the new version does the same job as `foreach [a b c][ add_all → a + b + c]`. If the models haven't converted to a new version it can be viewable as a read-only model but can't be executed. Other geospatial packages in R such as `rgdal` and `sf`, have also struggled whenever a major update was made on their own packages or on the R version itself due to a lot of dependencies. Even *ArcGIS*,

a UI (User Interface) software, had issues when they upgraded it from version 9.3 to 10. The projects that were coded under the VBA script in 9.3 were broken because it was not recognised as a correct function in the new version based on Python. This is also another example that backward compatibility and deprecation mechanisms are important.

Lastly, for more advanced users, it is also recommended to use a collaborative platform that executes every result from the codes with the exact version. One of the platforms is Codeocean (https://codeocean.com/). The Nature research team has recently chosen the platform to peer-review the codes (Perkel, 2019). The Nature editors and peer-reviewers strongly believed that coding has become a norm across many disciplines, and hence have asserted that the model process including the quality of data, conciseness, reproducibility, and documentation of the model should be placed as a requirement. Although the training procedure can be difficult at first, it will lead researchers to conduct themselves with more responsibility.

6.6 General Conclusion

To estimate population exposure to particulates from an individual level, this thesis applied ABM to improve on some aspects of the spatial and temporal dynamics from previous studies. To the author's knowledge, this thesis is one of the earliest studies to contextualise the framework of individual activities, exposure, and pollution, along with the demographic sensitivity that affects health outcomes.

The outcomes are summarised as follows:

- Chapter 3: Comparison of Spatial Interpolation Methods
  - While NO_2 showed a remarkable contrast between roadside and background as well as a daily cycle, PM_{10} had a small variance in daily cycles but had greater seasonal oscillation: populations may be exposed to high levels of PM_{10} by season regardless of location but that of NO_2 might be more variable.
  - Neither of the interpolation methods was noticeably superior to the other.
- Chapter 4: Assessing PM_{10} Exposure of Demographic ABM in Seoul Districts
  - The first agent model showed that disparities in health may arise as a result of differences in socioeconomic status, especially when the group was exposed over a long period
  - Road proximity is expected to cause additional health loss.
- Chapter 5: Estimating Exposure to PM_{10} with Agent-Based Traffic Simulations
  - Around 40% of the roadside PM_{10} can be contributed by non-exhaust emissions
  - Although local drivers were exposed to higher extreme values, longer exposure for pedestrians led to a larger accumulated exposure overall

Concluding remarks

Considering the complexity and variety of Seoul, even in the absence of good data linking exposure to actual health effects, some progress can be made using simulation. A much higher spatial density of monitoring data is required to make

progress in characterising exposure however, especially at extremes, and preferably at the level of the individual. Nevertheless, pollution exposure levels in a city can vary by both commuting patterns and also by the urban development of one's location. This work supports previous suggestions that people living in poorer areas may be more exposed to high atmospheric pollution loading. Local traffic levels make a significant contribution to this load. Even with vehicles that have low exhaust emissions, exposure to particulates from brake and tyre wear is likely to continue to be significant and may require traffic reduction policies; local modification of behaviour by individuals to try to avoid polluted areas during routine commuting may not be particularly helpful. The consequences may be explicitly useful for healthcare policy: in particular, aiding the most vulnerable groups and districts.

Bibliography

Aalto, J., Pirinen, P., Heikkinen, J., and Venäläinen, A. (2013). Spatial interpolation of monthly climate data for Finland: comparing the performance of kriging and generalized additive models. *Theoretical and Applied Climatology*, 112(1-2):99–111.

Air Quality Expert Group (2012). Fine Particulate Matter (PM2.5) in the United Kingdom. Technical report, Department for Environment, Food & Rural Affairs.

Air Quality Expert Group (2019). Non-Exhaust Emissions from Road Traffic. Technical report, Department for Environment, Food and Rural Affairs.

Al-Thani, H., Koç, M., Fountoukis, C., and Isaifan, R. J. (2020). Evaluation of particulate matter emissions from non-passenger diesel vehicles in Qatar. *Journal of the Air and Waste Management Association*, 70(2):228–242.

Altuğ, H., Fuks, K. B., Hüls, A., Mayer, A.-K., Tham, R., Krutmann, J., and Schikowski, T. (2020). Air pollution is associated with depressive symptoms in elderly women with cognitive impairment. *Environment International*, 136:105448.

Alves, C. A., Vicente, E. D., Vicente, A. M. P., Rienda, I. C., Tomé, M., Querol, X., and Amato, F. (2020). Loadings, chemical patterns and risks of inhalable road dust particles in an Atlantic city in the north of Portugal. *Science of The Total Environment*, 737:139596.

Amato, F. (2018). *Non-exhaust emissions: an urban air quality problem for public health; impact and mitigation measures*. Academic Press.

Amato, F., Cassee, F. R., Denier van der Gon, H. A. C., Gehrig, R., Gustafsson, M., Hafner, W., Harrison, R. M., Jozwicka, M., Kelly, F. J., Moreno, T., Prevot, A. S. H., Schaap, M., Sunyer, J., and Querol, X. (2014). Urban air quality: The challenge of traffic non-exhaust emissions. *Journal of Hazardous Materials*, 275:31–36.

An, L. (2012). Modeling human decisions in coupled human and natural systems: Review of agent-based models. *Ecological Modelling*, 229:25–36.

Aristodemou, E., Boganegra, L. M., Mottet, L., Pavlidis, D., Constantinou, A., Pain, C., Robins, A., and ApSimon, H. (2018). How tall buildings affect turbulent air flows and dispersion of pollution within a neighbourhood. *Environmental Pollution*, 233:782–796.

Atkinson, R. W., Analitis, A., Samoli, E., Fuller, G. W., Green, D. C., Mudway, I. S., Anderson, H. R., and Kelly, F. J. (2016). Short-term exposure to traffic-related air pollution and daily mortality in London, UK. *Journal of Exposure Science and Environmental Epidemiology*, 26(2):125–132.

Auchincloss, A. H. and Diez Roux, A. V. (2008). A new tool for epidemiology: The usefulness of dynamic-agent models in understanding place effects on health. *American Journal of Epidemiology*, 168(1):1–8.

Badham, J., Chattoe-Brown, E., Gilbert, N., Chalabi, Z., Kee, F., and Hunter, R. F. (2018). Developing agent-based models of complex health behaviour. *Health & Place*, 54:170–177.

Bae, C., Kim, B.-U., Kim, H. C., Yoo, C., and Kim, S. (2020). Long-range transport influence on key chemical components of pm2.5 in the seoul metropolitan area, south korea, during the years 2012–2016. *Atmosphere*, 11(1).

Bárdossy, A. (1997). Introduction to geostatistics. *Institute of Hydraulic Engineering, University of Stuttgart*.

Barratt, R. (2013). *Atmospheric dispersion modelling: an introduction to practical applications*. Routledge.

Batterman, S., Ganguly, R., and Harbin, P. (2015). High resolution spatial and temporal mapping of traffic-related air pollutants. *International Journal of Environmental Research and Public Health*, 12(4):3646–3666.

Bauduin, S., McIntire, E. J. B., and Chubaty, A. M. (2019). NetLogoR: a package to build and run spatially explicit agent-based models in R. *Ecography*, 42(11):1841–1849.

Baxter, L. K., Dionisio, K. L., Burke, J., Ebelt Sarnat, S., Sarnat, J. A., Hodas, N., Rich, D. Q., Turpin, B. J., Jones, R. R., Mannshardt, E., Kumar, N., Beevers, S. D., and Özkaynak, H. (2013). Exposure prediction approaches used in air pollution epidemiology studies: key findings and future recommendations. *Journal of exposure science & environmental epidemiology*, 23(6):654–659.

Beelen, R., Raaschou-Nielsen, O., Stafoggia, M., Andersen, Z. J., Weinmayr, G., Hoffmann, B., Wolf, K., Samoli, E., Fischer, P., Nieuwenhuijsen, M., Vineis, P., Xun, W. W., Katsouyanni, K., Dimakopoulou, K., Oudin, A., Forsberg, B., Modig, L., Havulinna, A. S., Lanki, T., Turunen, A., Oftedal, B., Nystad, W., Nafstad, P., De Faire, U., Pedersen, N. L., Östenson, C. G., Fratiglioni, L., Penell, J., Korek, M., Pershagen, G., Eriksen, K. T., Overvad, K., Ellermann, T., Eeftens, M., Peeters, P. H., Meliefste, K., Wang, M., Bueno-De-Mesquita, B., Sugiri, D., Krämer, U., Heinrich, J., De Hoogh, K., Key, T., Peters, A., Hampel, R., Concin, H., Nagel, G., Ineichen, A., Schaffner, E., Probst-Hensch, N., Künzli, N., Schindler, C., Schikowski, T., Adam, M., Phuleria, H., Vilier, A., Clavel-Chapelon, F., Declercq, C., Grioni, S., Krogh, V., Tsai, M. Y., Ricceri, F., Sacerdote, C., Galassi, C., Migliore, E., Ranzi, A., Cesaroni, G., Badaloni, C., Forastiere, F., Tamayo, I., Amiano, P., Dorronsoro, M., Katsoulis, M., Trichopoulou, A., Brunekreef, B., and Hoek, G. (2014). Effects of long-term exposure to air pollution on natural-cause mortality:

An analysis of 22 European cohorts within the multicentre ESCAPE project. *The Lancet*, 383(9919):785–795.

Beevers, S. D. (2016). Traffic management strategies for emissions reduction : recent experience in London. pages 27–39.

Beevers, S. D., Kitwiroon, N., Williams, M. L., Kelly, F. J., Ross Anderson, H., and Carslaw, D. C. (2013). Air pollution dispersion models for human exposure predictions in London. *Journal of exposure science & environmental epidemiology*, 23(6):647–53.

Bell, M. L. (2006). The use of ambient air quality modeling to estimate individual and population exposure for human health research: A case study of ozone in the Northern Georgia Region of the United States. *Environment International*, 32(5):586–593.

Benenson, I. and Torrens, P. (2004). *Geosimulation: Automata-based modeling of urban phenomena*. John Wiley & Sons.

Bhome, A. B. and Brashier, B. (2014). Profiles of chronic obstructive lung disease: characteristics of stable chronic obstructive lung disease in different parts of Asia. *Curr Opin Pulm Med*, 20(2):165–172.

Black, C., Tesfaigzi, Y., Bassein, J. A., and Miller, L. A. (2017). Wildfire smoke exposure and human health: Significant gaps in research for a growing public health issue. *Environmental Toxicology and Pharmacology*, 55:186–195.

Blanchard, C. L., Tanenbaum, S., and Hidy, G. M. (2014). Spatial and temporal variability of air pollution in Birmingham, Alabama. *Atmospheric Environment*, 89:382–391.

Bluett, J., Gimson, N., Fisher, G., Heydenrych, C., Freeman, T., and Godfrey, J. (2004). *Good practice guide for atmospheric dispersion modelling*.

Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences*, 99(suppl 3):7280 LP – 7287.

Borm, P. J. A., Kelly, F., Künzli, N., Schins, R. P. F., and Donaldson, K. (2007). Oxidant generation by particulate matter: from biologically effective dose to a promising, novel metric. *Occupational and environmental medicine*, 64(2):73–74.

Brauer, M., Freedman, G., Frostad, J., Van Donkelaar, A., Martin, R. V., Dentener, F., van Dingenen, R., Estep, K., Amini, H., and Apte, J. S. (2016). Ambient air pollution exposure estimation for the global burden of disease 2013. *Environmental science & technology*, 50(1):79–88.

Bravo, M. A., Fuentes, M., Zhang, Y., Burr, M. J., and Bell, M. L. (2012). Comparison of exposure estimation methods for air pollutants: ambient monitoring data and regional air quality simulation. *Environmental research*, 116:1–10.

Breuer, J. L., Samsun, R. C., Peters, R., and Stolten, D. (2020). The impact of diesel vehicles on NOx and PM10 emissions from road transport in urban morphological zones: A case study in North Rhine-Westphalia, Germany. *Science of The Total Environment*, 727:138583.

Briggs, D. J., Hoogh, K. D., Morris, C., and Gulliver, J. (2008). Effects of travel mode on exposures to particulate air pollution. 34:12–22.

Britannica, E. (2004). *Encyclopædia Britannica Online*. Britannica Online.

Brook, R. D., Franklin, B., Cascio, W., Hong, Y., Howard, G., Lipsett, M., Luepker, R., Mittleman, M., Samet, J., and Smith, S. C. (2004). Air pollution and cardiovascular disease A statement for healthcare professionals from the expert panel on population and prevention science of the American Heart Association. *Circulation*, 109(21):2655–2671.

Burrough, P. A., McDonnell, R., McDonnell, R. A., and Lloyd, C. D. (2015). *Principles of geographical information systems*. Oxford university press.

Byun, D. and Schere, K. L. (2006). Review of the Governing Equations, Computational Algorithms, and Other Components of the Models-3 Community Multiscale Air Quality (CMAQ) Modeling System. *Applied Mechanics Reviews*, 59(2):51.

Camana, F. A. and Deutsch, C. V. (2020). The Nugget Effect. *Geostatistics Lessons. Retrieved from http://www. geostatisticslessons. com/lessons/nuggeteffect GeostatisticsLessons. com*.

Cambridge Insight (2017). Cambridge Transport and Health. Technical report, Cambridge Insight.

Cambridgeshire City Council (2016). Air Pollution in Cambridgeshire. pages 1–44.

Caplin, A., Ghandehari, M., Lim, C., Glimcher, P., and Thurston, G. (2019). Advancing environmental exposure assessment science to benefit society. *Nature Communications*, 10(1):1236.

Cariolet, J.-M., Colombert, M., Vuillet, M., and Diab, Y. (2018). Assessing the resilience of urban areas to traffic-related air pollution: Application in Greater Paris. *Science of The Total Environment*, 615:588–596.

Carruthers, D. J., Holroyd, R. J., Hunt, J. C. R., Weng, W. S., Robins, A. G., Apsley, D. D., Thompson, D. J., and Smith, F. B. (1994). UK-ADMS: A new approach to modelling dispersion in the earth's atmospheric boundary layer. *Journal of Wind Engineering and Industrial Aerodynamics*, 52(C):139–153.

Castell, N., Dauge, F. R., Schneider, P., Vogt, M., Lerner, U., Fishbain, B., Broday, D., and Bartonova, A. (2017). Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates? *Environment International*, 99:293–302.

Castle, C. J. E. and Crooks, A. T. (2006). Principles and concepts of agent-based modelling for developing geospatial simulations.

Checkoway, H., Pearce, N., and Kriebel, D. (2007). Selecting appropriate study designs to address specific research questions in occupational epidemiology. *Occupational and environmental medicine*, 64(9):633–638.

Chen, B. and Kan, H. (2008). Air pollution and population health: a global challenge. *Environmental health and preventive medicine*, 13(2):94–101.

Chen, L., Mengersen, K., and Tong, S. (2007). Spatiotemporal relationship between particle air pollution and respiratory emergency hospital admissions in Brisbane, Australia. *Science of the Total Environment*, 373(1):57–67.

Chen, M., Yang, J., Hu, L., Hossain, M. S., and Muhammad, G. (2018). Urban healthcare big data system based on crowdsourced and cloud-based air quality indicators. *IEEE Communications Magazine*, 56(11):14–20.

Chiusolo, M., Cadum, E., Stafoggia, M., Galassi, C., Berti, G., Faustini, A., Bisanti, L., Vigotti, M. A., Dessì, M. P., Cernigliaro, A., Mallone, S., Pacelli, B., Minerba, S., Simonato, L., Forastiere, F., and Group, E. C. (2011). Short-Term Effects of Nitrogen Dioxide on Mortality and Susceptibility Factors in 10 Italian Cities: The EpiAir Study. *Environmental health perspectives*, 119(9):1233–1238.

Chung, S. Y., Venkatramanan, S., Elzain, H. E., Selvam, S., and Prasanna, M. V. (2019). Chapter 4 - Supplement of Missing Data in Groundwater-Level Variations of Peak Type Using Geostatistical Methods. pages 33–41. Elsevier.

Coffee, N. T., Lockwood, T., Hugo, G., Paquet, C., Howard, N. J., and Daniel, M. (2013). Relative residential property value as a socio-economic status indicator for health research. *International journal of health geographics*, 12(1):22.

Collier, N. (2003). Repast: An extensible framework for agent simulation. *The University of Chicago's Social Science Research*, 36:2003.

Coupé, C. (2018). Modeling Linguistic Variables With Regression Models: Addressing Non-Gaussian Distributions, Non-independent Observations, and Non-linear Predictors With Random Effects and Generalized Additive Models for Location, Scale, and Shape. *Frontiers in Psychology*, 9:513.

Crooks, A., Malleson, N., Manley, E., and Heppenstall, A. (2018). *Agent-based modelling and geographical information systems: a practical primer*. SAGE Publications Limited.

Crooks, A., Malleson, N., Manley, E., and Heppenstall, A. (2019). *Agent-Based Modelling and Geographical Information Systems: A Practical Primer (Spatial Analytics and GIS)*. SAGE Publications Ltd, 1 edition.

Cyrys, J., Peters, A., Soentgen, J., and Wichmann, H.-E. (2014). Low emission zones reduce PM10 mass concentrations and diesel soot in German cities. *Journal of the Air & Waste Management Association*, 64(4):481–487.

David, N. and Don, G. (2012). An integrated agent-based framework for assessing air pollution impacts. *Journal of Environmental Protection*, 2012.

DEFRA (2003). Conversion Factors Between ppb and µg m-3 and ppm and mgm-3.

Deligiorgi, D. and Philippopoulos, K. (2011). Spatial Interpolation Methodologies in Urban Air Pollution Modeling: Application for the Greater Area of Metropolitan Athens, Greece. In Nejadkoorki, F., editor, *Advanced Air Pollution*. InTech.

Deshpande, S. S. (2002). *Handbook of food toxicology*. CRC Press.

Dewulf, B., Neutens, T., Lefebvre, W., Seynaeve, G., Vanpoucke, C., Beckx, C., and Van de Weghe, N. (2016). Dynamic assessment of exposure to air pollution using mobile phone data. *International Journal of Health Geographics*, 15(1):1.

Di Sabatino, S., Buccolieri, R., Pulvirenti, B., and Britter, R. E. (2008). Flow and pollutant dispersion in street canyons using FLUENT and ADMS-Urban. *Environmental Modeling & Assessment*, 13(3):369–381.

Dias, D. and Tchepel, O. (2018). Spatial and Temporal Dynamics in Air Pollution Exposure Assessment. *International journal of environmental research and public health*, 15(3):558.

Dong, G.-H., Chen, T., Liu, M.-M., Wang, D., Ma, Y.-N., Ren, W.-H., Lee, Y. L., Zhao, Y.-D., and He, Q.-C. (2011). Gender differences and effect of air pollution on asthma in children with and without allergic predisposition: northeast Chinese children health study. *PloS one*, 6(7):e22470.

Dons, E., Kochan, B., Bellemans, T., Wets, G., and Panis, L. I. (2014). Modeling Personal Exposure to Air Pollution with AB2C: Environmental Inequality. *Procedia Computer Science*, 32:269–276.

Dunham, J. B. (2005). An agent-based spatially explicit epidemiological model in MASON. *Journal of Artificial Societies and Social Simulation*, 9(1).

Edmonds, B., Le Page, C., Bithell, M., Chattoe-Brown, E., Grimm, V., Meyer, R., Montañola-Sales, C., Ormerod, P., Root, H., and Squazzoni, F. (2019). Different Modelling Purposes. *Journal of Artificial Societies and Social Simulation*, 22(3):6.

Einstein, A. (2010). Everything should be made as simple as possible, but not simpler.

EMEP/EEA (2019). EMEP/EEA Air Pollutant Emission Inventory Guidebook 2019. Technical report, European Environment Agency.

Epstein, J. M. (2008). Why model? *Journal of Artificial Societies and Social Simulation*, 11(4):12.

Europe, W. (2013). Proximity to roads, NO2, other air pollutants and their mixtures. In *Review of evidence on health aspects of air pollution–REVIHAAP Project: Technical Report [Internet]*. WHO Regional Office for Europe.

European Commission (2014). Recommendation from the Scientific Committee on Occupational Exposure Limits for Nitrogen Dioxide. pages 1–17.

European Lung Foundation (2020). Walking quieter routes to work can avoid peaks in air pollution. *ScienceDaily*.

Fann, N., Lamson, A. D., Anenberg, S. C., Wesson, K., Risley, D., and Hubbell, B. J. (2012). Estimating the national public health burden associated with exposure to ambient PM2.5 and ozone. *Risk analysis : an official publication of the Society for Risk Analysis*, 32(1):81–95.

Farmer, J. D. and Foley, D. (2009). The economy needs agent-based modelling. *Nature*, 460(7256):685–686.

Felsen, M. and Wilensky, U. (2007). NetLogo Urban Suite - Pollution model.

Ferm, M. and Sjöberg, K. (2015). Concentrations and emission factors for PM2.5 and PM10 from road traffic in Sweden. *Atmospheric Environment*, 119:211–219.

Fowler, H. J. and Ekström, M. (2009). Multi-model ensemble estimates of climate change impacts on UK seasonal rainfall extremes. *Int. J. Climatol*, 416(January):385–416.

Friedland, C. J., Joyner, T. A., Massarra, C., Rohli, R. V., Treviño, A. M., Ghosh, S., Huyck, C., and Weatherhead, M. (2017). Isotropic and anisotropic kriging approaches for interpolating surface-level wind speeds across large, geographically diverse regions. *Geomatics, Natural Hazards and Risk*, 8(2):207–224.

Fritz, J., Neuweiler, I., and Nowak, W. (2009). Application of FFT-based Algorithms for Large-Scale Universal Kriging Problems. *Mathematical Geosciences*, 41(5):509–533.

Galea, S., Riddle, M., and Kaplan, G. A. (2010). Causal thinking and complex system approaches in epidemiology. *International Journal of Epidemiology*, 39(1):97–106.

G.D., T., H., K., I., A., J., B., R.D., B., K., C., S., D. M., F., F., B., F., M.W., F., J., G., D., H., F.J., K., N., K., R., L., A., P., S.T., R., D., R., B., R., J.M., S., T., S., T., S., J., S., and Brunekreef, B. (2017). A joint ERS/ATS policy statement: What constitutes an adverse health effect of air pollution? An analytical framework. *European Respiratory Journal*, 49(1) (pagination):Arte Number: 1600419. ate of Pubaton: 01 Jan 2017.

Gilbert, N. (2008). Using Agent-Based Models in Social Science Research. *Agent-Based Models*, pages 30–46.

González, E. J. (2012). *Artificial Intelligence Resources in Control and Automation Engineering*. Bentham Science Publishers.

Greater London Authority (2017). London Atmospheric Emissions Inventory (LAEI) 2013. *London: London Datastore*.

Grimm, V., Berger, U., DeAngelis, D. L., Polhill, J. G., Giske, J., and Railsback, S. F. (2010). The ODD protocol: a review and first update. *Ecological modelling*, 221(23):2760–2768.

Grimm, V., Railsback, S. F., Vincenot, C. E., Berger, U., Gallagher, C., DeAngelis, D. L., Edmonds, B., Ge, J., Giske, J., Groeneveld, J., Johnston, A. S. A., Milles, A., Nabe-Nielsen, J., Polhill, J. G., Radchuk, V., Rohwäder, M.-S., Stillman, R. A., Thiele, J. C., and Ayllón, D. (2020). The ODD Protocol for Describing Agent-Based and Other Simulation Models: A Second Update to Improve Clarity, Replication, and Structural Realism. *Journal of Artificial Societies and Social Simulation*, 23(2):7.

Guan, D., Su, X., Zhang, Q., Peters, G. P., Liu, Z., Lei, Y., and He, K. (2014). The socioeconomic drivers of China's primary PM 2.5 emissions. *Environmental Research Letters*, 9(2):024010.

Guarnieri, M. and Balmes, J. R. (2014). Outdoor air pollution and asthma. *Lancet*, 383(9928):1581–1592.

Gulliver, J. and Briggs, D. J. (2007). Journey-time exposure to particulate air pollution. *Atmospheric Environment*, 41(34):7195–7207.

Gurram, S., Stuart, A. L., and Pinjari, A. R. (2015). Impacts of travel activity and urbanicity on exposures to ambient oxides of nitrogen and on exposure disparities. *Air Quality, Atmosphere & Health*, 8(1):97–114.

Gurram, S., Stuart, A. L., and Pinjari, A. R. (2019). Agent-based modeling to estimate exposures to urban air pollution from transportation: Exposure disparities and impacts of high-resolution data. *Computers, Environment and Urban Systems*, 75(April 2018):22–34.

Hafenbrädl, S., Waeger, D., Marewski, J. N., and Gigerenzer, G. (2016). Applied decision making with fast-and-frugal heuristics. *Journal of Applied Research in Memory and Cognition*, 5(2):215–231.

Halonen, J. I., Blangiardo, M., Toledano, M. B., Fecht, D., Gulliver, J., Anderson, H. R., Beevers, S. D., Dajnak, D., Kelly, F. J., and Tonne, C. (2016). Long-term exposure to traffic pollution and hospital admissions in London. *Environmental Pollution*, 208:48–57.

Hammond, R. A. (2015). Considerations and best practices in agent-based modeling to inform policy. In *Assessing the use of agent-based models for tobacco regulation*. National Academies Press (US).

Hamra, G. B., Guha, N., Cohen, A., Laden, F., Raaschou-Nielsen, O., Samet, J. M., Vineis, P., Forastiere, F., Saldiva, P., and Yorifuji, T. (2014). Outdoor particulate matter exposure and lung cancer: a systematic review and meta-analysis. *Environmental health perspectives*.

Hancock, D. (2020). Can air pollution be fatal for children with asthma? *Journal of Health Visiting*, 8(2):70–73.

Hankey, S. and Marshall, J. D. (2015). Land Use Regression Models of On-Road Particulate Air Pollution (Particle Number, Black Carbon, PM2.5, Particle Size) Using Mobile Monitoring. *Environmental Science & Technology*, 49(15):9194–9202.

Harrison, R. M., Jones, A. M., and Lawrence, R. G. (2004). Major component composition of PM10and PM2.5from roadside and urban background sites. *Atmospheric Environment*, 38(27):4531–4538.

He, M. Z., Kinney, P. L., Li, T., Chen, C., Sun, Q., Ban, J., Wang, J., Liu, S., Goldsmith, J., and Kioumourtzoglou, M.-A. (2020). Short- and intermediate-term exposure to NO2 and mortality: A multi-county analysis in China. *Environmental Pollution*, 261:114165.

Heinrich, J., Thiering, E., Rzehak, P., Krämer, U., Hochadel, M., Rauchfuss, K. M., Gehring, U., and Wichmann, H. E. (2013). Long-term exposure to NO 2 and PM 10 and all-cause and cause-specific mortality in a prospective cohort of women. *Occupational and Environmental Medicine*, 70(3):179–186.

Hoek, G., Beelen, R., de Hoogh, K., Vienneau, D., Gulliver, J., Fischer, P., and Briggs, D. (2008). A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmospheric Environment*, 42(33):7561–7578.

Hoek, G., Krishnan, R. M., Beelen, R., Peters, A., Ostro, B., Brunekreef, B., and Kaufman, J. D. (2013). Long-term air pollution exposure and cardio-respiratory mortality: a review. *Environmental Health*, 12(1):43.

Holgate, S. T. (2017). 'Every breath we take: the lifelong impact of air pollution'–a call for action. *Clinical Medicine*, 17(1):8.

Holland, J. (2003). Applying complexity science to health and healthcare. *Minneapolis (MN): Center for the Study of Health care Management, University of Minnesota.*

Holmes, N. S. and Morawska, L. (2006). A review of dispersion modelling and its application to the dispersion of particles: An overview of different dispersion models available. *Atmospheric Environment*, 40(30):5902–5928.

Huang, H., Akustu, Y., Arai, M., and Tamura, M. (2001). Analysis of photochemical pollution in summer and winter using a photochemical box model in the center of Tokyo, Japan. *Chemosphere*, 44(2):223–230.

Hülsmann, F., Gerike, R., and Ketzel, M. (2014). Modelling traffic and air pollution in an integrated approach - the case of Munich. *Urban Climate*, 10:732–744.

Hülsmann, F., Gerike, R., Kickhöfer, B., Nagel, K., and Luz, R. (2011). Towards a multi-agent based modeling approach for air pollutants in urban regions. *Luftqualität an Straßen*, (March 2016):1–16.

Hurley, P. (2002). *The Air Pollution Model (TAPM) Version 2: User Manual.* CSIRO Atmospheric Research.

Hurley, P. (2008). TAPM V4. User manual. *CSIRO Marine and Atmospheric Research Internal Report No. 5.*

Hwang, Y. and Lee, K. (2018). Contribution of microenvironments to personal exposures to PM10 and PM2.5 in summer and winter. *Atmospheric Environment*, 175:192–198.

IARC (2013). *IARC: Outdoor air pollution a leading environmental cause of cancer deaths*. International Agency for Research on Cancer.

Idarraga, M. A., Guerrero, J. S., Mosle, S. G., Miralles, F., Galor, A., and Kumar, N. (2020). Relationships Between Short-Term Exposure to an Indoor Environment and Dry Eye (DE) Symptoms. *Journal of clinical medicine*, 9(5):1316.

Jaxa-Rozen, M. and Kwakkel, J. H. (2018). Pynetlogo: Linking netlogo with python. *Journal of Artificial Societies and Social Simulation*, 21(2).

Jerrett, M., Burnett, R. T., Kanaroglou, P., Eyles, J., Finkelstein, N., Giovis, C., and Brook, J. R. (2001). A GIS - Environmental justice analysis of particulate air pollution in Hamilton, Canada. *Environment and Planning A*, 33(6):955–973.

Jerrett, M., Gale, S., and Kontgis, C. (2009). An environmental health geography of risk. In *A Companion to Health and Medical Geography*.

Jerrett, M., Gale, S., and Kontgis, C. (2010). Spatial modeling in environmental and public health research. *International Journal of Environmental Research and Public Health*, 7(4):1302–1329.

Jung, Y., Jeong, K. H.-s., Ma, S. K. Y.-i., Jeongsoo, W.-k. L., and Young, K. (2016). Seasonal nitrogen oxides improvement due to on-road mobile air pollution source emission control plan in seoul metropolitan area. 38(5):269–278.

Kan, H. and Chen, B. (2004). Particulate air pollution in urban areas of Shanghai, China: Health-based economic assessment. *Science of the Total Environment*, 322(1-3):71–79.

Kan, H., London, S. J., Chen, G., Zhang, Y., Song, G., Zhao, N., Jiang, L., and Chen, B. (2008). Season, sex, age, and education as modifiers of the effects of outdoor air pollution on daily mortality in Shanghai, China: The Public Health and Air Pollution in Asia (PAPA) Study. *Environ Health Perspect*, 116(9):1183–1188.

Kaur, S., Nieuwenhuijsen, M. J., and Colvile, R. N. (2005). Pedestrian exposure to air pollution along a major road in Central London, UK. *Atmospheric Environment*, 39(38):7307–7320.

Kelly, F., Armstrong, B., Atkinson, R., Anderson, H. R., Barratt, B., Beevers, S., Cook, D., Green, D., Derwent, D., Mudway, I., Wilkinson, P., and Committee, H. H. R. (2011). The London low emission zone baseline study. *Research report (Health Effects Institute)*, (163):3–79.

Keuken, M. P., Henzing, J. S., Zandveld, P., van den Elshout, S., and Karl, M. (2012). Dispersion of particle numbers and elemental carbon from road traffic, a harbour and an airstrip in the Netherlands. *Atmospheric Environment*, 54(Complete):320–327.

Khajeh-Hosseini-Dalasm, N. and Longest, P. W. (2015). Deposition of particles in the alveolar airways: inhalation and breath-hold with pharmaceutical aerosols. *Journal of aerosol science*, 79:15–30.

Kim, D.-s., Jeong, J., and Ahn, J. (2016). Characteristics in Atmospheric Chemistry between NO , NO 2 and O 3 at an Urban Site during MAPS (Megacity Air Pollution Study) -Seoul , Korea. (2):422–434.

Kim, J. J., Smorodinsky, S., Lipsett, M., Singer, B. C., Hodgson, A. T., and Ostro, B. (2004). Traffic-related air pollution near busy roads: the East Bay Children's Respiratory Health Study. *American journal of respiratory and critical care medicine*, 170(5):520–526.

Kim, S.-Y., Yi, S.-J., Eum, Y. S., Choi, H.-J., Shin, H., Ryou, H. G., and Kim, H. (2014). Ordinary kriging approach to predicting long-term particulate matter concentrations in seven major Korean cities. *Environmental Health and Toxicology*, 29:e2014012.

Kiš, I. M. (2016). Comparison of Ordinary and Universal Kriging interpolation techniques on a depth variable (a case of linear spatial trend), case study of the Šandrovac Field. *Rudarsko-geološko-naftni zbornik*, 31(2):41–58.

Knowlton, K., Rosenthal, J. E., Hogrefe, C., Lynn, B., Gaffin, S., Goldberg, R., Rosenzweig, C., Civerolo, K., Ku, J.-y., Kinney, P. L., and Rosenthal, E. (2004). Assessing Ozone-Related Health Impacts Under a Chaning Climate. *Environmental Health Perspectives*, 112(15):1557–1563.

Kobza, J. and Geremek, M. (2017). Do the pollution related to high-traffic roads in urbanised areas pose a significant threat to the local population? *Environmental Monitoring and Assessment*, 189(1).

Koch, A. and Carson, D. (2012). Spatial, Temporal and Social Scaling in Sparsely Populated Areas Geospatial Mapping and Simulation Techniques to Investigate Social Diversity. *GI_Forum*, pages 44–53.

Korner-Nievergelt, F., Roth, T., Von Felten, S., Guélat, J., Almasi, B., and Korner-Nievergelt, P. (2015). *Bayesian data analysis in ecology using linear models with R, BUGS, and Stan*. Academic Press.

Kreider, M. L., Unice, K. M., and Panko, J. M. (2019). Human health risk assessment of Tire and Road Wear Particles (TRWP) in air. *Human and Ecological Risk Assessment*, 0(0):1–19.

Kumar, P., Hama, S., Nogueira, T., Abbass, R. A., Brand, V. S., de Fatima Andrade, M., Asfaw, A., Aziz, K. H., Cao, S.-J., El-Gendy, A., Islam, S., Jeba, F., Khare, M., Mamuya, S. H., Martinez, J., Meng, M.-R., Morawska, L., Muula, A. S., S M, S. N., Ngowi, A. V., Omer, K., Olaya, Y., Osano, P., and Salam, A. (2020). In-car particulate matter exposure across ten global cities. *Science of The Total Environment*, 750:141395.

Kumar, V. (2007). Optimal contour mapping of groundwater levels using universal kriging - A case study. *Hydrological Sciences Journal*, 52(5):1038–1050.

Kwak, J. H., Kim, H., Lee, J., and Lee, S. (2013). Characterization of non-exhaust coarse and fine particles from on-road driving and laboratory measurements. *Science of the Total Environment*, 458-460:273–282.

Kwon, S.-B., Jeong, W., Park, D., Kim, K.-T., and Cho, K. H. (2015). A multivariate study for characterizing particulate matter (PM10, PM2.5, and PM1) in Seoul metropolitan subway stations, Korea. *Journal of Hazardous Materials*, 297:295–303.

Langrish, J. P., Mills, N. L., Chan, J. K. K., Leseman, D. L. A. C., Aitken, R. J., Fokkens, P. H. B., Cassee, F. R., Li, J., Donaldson, K., and Newby, D. E. (2009). Beneficial cardiovascular effects of reducing exposure to particulate air pollution with a simple facemask. *Particle and fibre toxicology*, 6(1):8.

Lark, R. M. (2009). Kriging a soil variable with a simple nonstationary variance model. *Journal of agricultural, biological, and environmental statistics*, 14(3):301–321.

Larkin, A. and Hystad, P. (2017). Towards Personal Exposures: How Technology Is Changing Air Pollution and Health Research. *Current environmental health reports*, 4(4):463–471.

Laumbach, R., Meng, Q., and Kipen, H. (2015). What can individuals do to reduce personal health risks from air pollution? *Journal of thoracic disease*, 7(1):96–107.

Law, M. and Collins, A. (2019). *Getting to know ArcGIS PRO*. Esri press.

Lawrence, S., Sokhi, R., Ravindra, K., Mao, H., Prain, H. D., and Bull, I. D. (2013). Source apportionment of traffic emissions of particulate matter using tunnel measurements. *Atmospheric Environment*, 77:548–557.

Lee, Hwang, and Kim (2014). Health Vulnerability Assessment for PM10 in Busan. *The Journal of Korean Environmental Sciences Society*, 40(5):355–366.

Lee, B.-J. and Park, S.-S. (2019). Evaluation of PM10 and PM2.5 Concentrations from Online Light Scattering Dust Monitors Using Gravimetric and Beta-ray Absorption Methods. *Journal of Korean Society for Atmospheric Environment*, 35(3):357–369.

Lee, H. M. and Kim, Y. P. (2007). Analysis on the Effects of Traffic Control Program on the Air Quality in Seoul. 23(4):498–506.

Lee, M., Brauer, M., Wong, P., Tang, R., Tsui, T. H., Choi, C., Cheng, W., Lai, P.-C., Tian, L., Thach, T.-Q., Allen, R., and Barratt, B. (2017). Land use regression modelling of air pollution in high density high rise cities: A case study in Hong Kong. *Science of The Total Environment*, 592:306–315.

Lemos, L. L. and Pasin, M. (2016). Intersection control in transportation networks: Opportunities to minimize air pollution emissions. In *Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International Conference on*, pages 1616–1621. IEEE.

Leung, D. Y. C. (2015). Outdoor-indoor air pollution in urban environment: challenges and opportunity. *Frontiers in Environmental Science*, 2:69.

Li, J. and Heap, A. D. (2014). Spatial interpolation methods applied in the environmental sciences: A review. *Environmental Modelling & Software*, 53:173–189.

Li, L., Wu, J., Wilhelm, M., and Ritz, B. (2012). Use of generalized additive models and cokriging of spatial residuals to improve land-use regression estimates of nitrogen oxides in Southern California. *Atmospheric Environment*, 55:220–228.

Liang, L., Gong, P., Cong, N., Li, Z., Zhao, Y., and Chen, Y. (2019). Assessment of personal exposure to particulate air pollution: the first result of City Health Outlook (CHO) project. *BMC Public Health*, 19(1):711.

Liu, J., Li, J., and Li, W. (2016). Temporal patterns in fine particulate matter time series in beijing: A calendar view. *Scientific Reports*, 6(1):32221.

Loo, B. P. Y., Lam, W. W. Y., Mahendran, R., and Katagiri, K. (2017). How Is the Neighborhood Environment Related to the Health of Seniors Living in Hong Kong, Singapore, and Tokyo? Some Insights for Promoting Aging in Place. *Annals of the American Association of Geographers*, pages 1–17.

Loomis, D., Grosse, Y., Lauby-Secretan, B., El Ghissassi, F., Bouvard, V., Benbrahim-Tallaa, L., Guha, N., Baan, R., Mattock, H., and Straif, K. (2013). The carcinogenicity of outdoor air pollution. *Lancet Oncology*, 14(13):1262.

Luo, W., Taylor, M. C., and Parker, S. R. (2008). A comparison of spatial interpolation methods to estimate continuous wind speed surfaces using irregularly distributed data from England and Wales. *International journal of climatology*, 28(7):947–959.

Maantay, J. (2007). Asthma and air pollution in the Bronx: methodological and data considerations in using GIS for environmental justice and health research. *Health & place*, 13(1):32–56.

Malvić, T. and Balić, D. (2009). Linearity and Lagrange Linear Multiplicator in the Equations of Ordinary Kriging. *Nafta: exploration, production, processing, petrochemistry*, 60(1):31–43.

Manley, E. and Cheng, T. (2018). Exploring the role of spatial cognition in predicting urban traffic flow through agent-based modelling. *Transportation Research Part A: Policy and Practice*, 109:14–23.

Manley, E. J., Addison, J. D., and Cheng, T. (2015a). Shortest path or anchor-based route choice: a large-scale empirical analysis of minicab routing in London. *Journal of Transport Geography*, 43:123–139.

Manley, E. J., Orr, S. W., and Cheng, T. (2015b). A heuristic model of bounded route choice in urban areas. *Transportation Research Part C: Emerging Technologies*, 56:195–209.

Mao, P., Li, J., Xiong, L., Wang, R., Wang, X., Tan, Y., and Li, H. (2019). Characterization of Urban Subway Microenvironment Exposure- A Case of Nanjing in China. *International journal of environmental research and public health*, 16(4):625.

Masad, D. and Kazil, J. (2015). MESA: an agent-based modeling framework. In *14th PYTHON in Science Conference*, pages 53–60. Citeseer.

Mastio, M., Zargayouna, M., Scemama, G., and Rana, O. (2018). Distributed agent-based traffic simulations. *IEEE Intelligent Transportation Systems Magazine*, 10(1):145–156.

Miller, K. A., Siscovick, D. S., Sheppard, L., Shepherd, K., Sullivan, J. H., Anderson, G. L., and Kaufman, J. D. (2007). Long-Term Exposure to Air Pollution and Incidence of Cardiovascular Events in Women. *New England Journal of Medicine*, 356(5):447–458.

Min, K.-D., Yi, S.-J., Kim, H.-C., Leem, J.-H., Kwon, H.-J., Hong, S., Kim, K. S., and Kim, S.-Y. (2020). Association between exposure to traffic-related air pollution and pediatric allergic diseases based on modeled air pollution concentrations and traffic measures in Seoul, Korea: a comparative analysis. *Environmental Health*, 19(1):6.

Mitchell, G. and Dorling, D. (2003). An environmental justice analysis of British air quality. *Environment and Planning A*, 35(5):909–929.

Moreno-Jiménez, A., Cañada-Torrecilla, R., Vidal-Domínguez, M. J., Palacios-García, A., and Martínez-Suárez, P. (2016). Assessing environmental justice through potential exposure to air pollution: A socio-spatial analysis in Madrid and Barcelona, Spain. *Geoforum*, 69:117–131.

Moritz, S. (2019). Package 'imputeTS'. *cran. r-project. org*.

Moritz, S. and Bartz-Beielstein, T. (2017). imputeTS: time series missing value imputation in R. *The R Journal*, 9(1):207–218.

Moshammer, H., Hutter, H.-P., Hauck, H., and Neuberger, M. (2006). Low levels of air pollution induce changes of lung function in a panel of schoolchildren. *European Respiratory Journal*, 27(6):1138 LP – 1143.

Mudway, I. S., Dundas, I., Wood, H. E., Marlin, N., Jamaludin, J. B., Bremner, S. A., Cross, L., Grieve, A., Nanzer, A., Barratt, B. M., Beevers, S., Dajnak, D., Fuller, G. W., Font, A., Colligan, G., Sheikh, A., Walton, R., Grigg, J., Kelly, F. J., Lee, T. H., and Griffiths, C. J. (2019). Impact of London's low emission zone on air quality and children's respiratory health: a sequential annual cross-sectional study. *The Lancet Public Health*, 4(1):e28–e40.

Munir, S., Mayfield, M., Coca, D., and Jubb, S. A. (2019). Structuring an integrated air quality monitoring network in large urban areas – Discussing the purpose, criteria and deployment strategy. *Atmospheric Environment: X*, 2:100027.

Naughton, O., Donnelly, A., Nolan, P., Pilla, F., Misstear, B. D., and Broderick, B. (2018). A land use regression model for explaining spatial variation in air pollution levels using a wind sector based approach. *Science of The Total Environment*, 630:1324–1334.

Nikolova, I., Janssen, S., Vos, P., and Berghmans, P. (2014). Modelling the mixing of size resolved traffic induced and background ultrafine particles from an urban street canyon to adjacent backyards. *Aerosol Air Qual. Res*, 14(1):145–155.

Norton, J. (2015). An introduction to sensitivity assessment of simulation models. *Environmental Modelling & Software*, 69:166–174.

Nyhan, M., Grauwin, S., Britter, R., Misstear, B., McNabola, A., Laden, F., Barrett, S. R. H., and Ratti, C. (2016). "exposure track" - The impact of mobile-device-based mobility patterns on quantifying population exposure to air pollution. *Environmental Science and Technology*, 50(17):9671–9681.

Oanh, N. T. K. (2012). *Integrated Air Quality Management: Asian Case Studies*. CRC Press.

Odekanle, E. L., Fakinle, B. S., Jimoda, L. A., Okedere, O. B., Akeredolu, F. A., and Sonibare, J. A. (2017). In-vehicle and pedestrian exposure to carbon monoxide and volatile organic compounds in a mega city. *Urban Climate*, 21:173–182.

of Seoul, G. (2016). Seoul Air Pollution Report 2016. Technical report.

O'Neill, M. S., Jerrett, M., Kawachi, I., Levy, J. I., Cohen, A. J., Gouveia, N., Wilkinson, P., Fletcher, T., Cifuentes, L., and Schwartz, J. (2003). Health, wealth, and air pollution: advancing theory and methods. *Environmental health perspectives*, 111(16):1861.

Ott, W. R., Steinemann, A. C., and Wallace, L. A. (2006). *Exposure analysis*. CRC Press.

Özkaynak, H., Baxter, L. K., Dionisio, K. L., and Burke, J. (2013). Air pollution exposure prediction approaches used in air pollution epidemiology studies. *Journal of exposure science & environmental epidemiology*, 23(6):566–572.

Panko, J., Chu, J., Kreider, M. L., and Unice, K. M. (2013). Measurement of airborne concentrations of tire and road wear particles in urban and rural areas of France, Japan, and the United States. *Atmospheric Environment*, 72:192–199.

Paramasivam, C. R. and Venkatramanan, S. (2019). Chapter 3 - An Introduction to Various Spatial Analysis Techniques. pages 23–30. Elsevier.

Park, Y.-S. and Lek, S. (2016). Chapter 7 - Artificial Neural Networks: Multilayer Perceptron for Ecological Modeling. In Jørgensen, S. E., editor, *Ecological Model Types*, volume 28 of *Developments in Environmental Modelling*, pages 123–140. Elsevier.

Parmentier, B., McGill, B., Wilson, A. M., Regetz, J., Jetz, W., Guralnick, R. P., Tuanmu, M. N., Robinson, N., and Schildhauer, M. (2014). An assessment of methods and remote-sensing derived covariates for regional predictions of 1 km daily maximum air temperature. *Remote Sensing*, 6(9):8639–8670.

Parry, M. L. (2007). *Climate Change 2007: Impacts, Adaptation and Vulnerability: Summary for Policymakers, a Report of Working Group II of the Intergovernmental Panel on Climate Change and Technical Summary*. Cambridge University Press.

Pasquill, F. (1961). The estimation of the dispersion of windborne material. *Meteorological magazine*, 90(1063).

Pearce, J., Kingham, S., and Zawar-Reza, P. (2006). Every breath you take? Environmental justice and air pollution in Christchurch, New Zealand. *Environment and Planning A*, 38(5):919–938.

Pearce, J. R., Richardson, E. A., Mitchell, R. J., and Shortt, N. K. (2011). Environmental justice and health: a study of multiple environmental deprivation and geographical inequalities in health in New Zealand. *Social Science & Medicine*, 73(3):410–420.

Pebesma, E. (2019). The meuse data set: a brief tutorial for the gstat R package.

Pebesma, E., Graeler, B., and Pebesma, M. E. (2019). Package 'gstat'.

Pedersen, E. J., Miller, D. L., Simpson, G. L., and Ross, N. (2019). Hierarchical generalized additive models in ecology: an introduction with mgcv.

Perkel, J. M. (2019). Make code accessible with these cloud services. *Nature*, 575(7781):247.

Perricone, G., Matějka, V., Alemani, M., Valota, G., Bonfanti, A., Ciotti, A., Olofsson, U., Söderberg, A., Wahlström, J., Nosko, O., Straffelini, G., Gialanella, S., and Ibrahim, M. (2018). A concept for reducing PM10 emissions for car brakes by 50%. *Wear*, 396-397:135–145.

Pershagen, G., Rylander, E., Norberg, S., Eriksson, M., and Nordvall, S. L. (1995). Air pollution involving nitrogen dioxide exposure and wheezing bronchitis in children. *International journal of epidemiology*, 24(6):1147–1153.

Petkus, A. J., Younan, D., Wang, X., Serre, M., Vizuete, W., Resnick, S., Espeland, M. A., Gatz, M., Chui, H., Manson, J. E., and Chen, J.-C. (2019). Particulate Air Pollutants and Trajectories of Depressive Symptoms in Older Women. *The American Journal of Geriatric Psychiatry*, 27(10):1083–1096.

Pettygrove, S. (2016). Dose-response relationship.

Railsback, S. F., Ayllón, D., Berger, U., Grimm, V., Lytinen, S., Sheppard, C., and Thiele, J. (2017). Improving Execution Speed of Models Implemented in NetLogo. *Journal of Artificial Societies and Social Simulation*, 20(1):3.

Railsback, S. F. and Grimm, V. (2011). *Agent-based and individual-based modeling: a practical introduction*. Princeton university press.

Reames, T. G. and Bravo, M. A. (2019). People, place and pollution: Investigating relationships between air quality perceptions, health concerns, exposure, and individual- and area-level characteristics. *Environment International*, 122:244–255.

Renaud, F. G. (2006). *Environmental components of vulnerability*. United Nations University Press, Tokyo, Japan.

Rivas, I., Kumar, P., and Hagen-Zanker, A. (2017). Exposure to air pollutants during commuting in London: Are there inequalities among different socio-economic groups? *Environment International*, 101:143–157.

Roberts–Semple, D., Song, F., and Gao, Y. (2012). Seasonal characteristics of ambient nitrogen oxides and ground–level ozone in metropolitan northeastern New Jersey. *Atmospheric Pollution Research*, 3(2):247–257.

Sacks, J. D., Stanek, L. W., Luben, T. J., Johns, D. O., Buckley, B. J., Brown, J. S., and Ross, M. (2011). Particulate matter-induced health effects: who is susceptible? *Environmental health perspectives*, 119(4):446.

Salecker, J., Sciaini, M., Meyer, K. M., and Wiegand, K. (2019). The nlrx r package: A next-generation framework for reproducible NetLogo model analyses. *Methods in Ecology and Evolution*, 10(11):1854–1863.

Samoli, E., Atkinson, R. W., Analitis, A., Fuller, G. W., Green, D. C., Mudway, I., Anderson, H. R., and Kelly, F. J. (2016). Associations of short-term exposure to traffic-related air pollution with cardiovascular and respiratory hospital admissions in London, UK. *Occupational and Environmental Medicine*, 73(5):300–307.

Sanchez, M., Milà, C., Sreekanth, V., Balakrishnan, K., Sambandam, S., Nieuwenhuijsen, M., Kinra, S., Marshall, J. D., and Tonne, C. (2020). Personal exposure to particulate matter in peri-urban India: predictors and association with ambient concentration at residence. *Journal of Exposure Science and Environmental Epidemiology*, 30(4):596–605.

Schafer, J. L. (1999). Multiple imputation: a primer. *Statistical methods in medical research*, 8(1):3–15.

Schneider, P., Castell, N., Vogt, M., Dauge, F. R., Lahoz, W. A., and Bartonova, A. (2017). Mapping urban air quality in near real-time using observations from low-cost sensors and model information. *Environment International*, 106:234–247.

Schraufnagel, D. E., Balmes, J. R., Cowl, C. T., De Matteis, S., Jung, S.-H., Mortimer, K., Perez-Padilla, R., Rice, M. B., Riojas-Rodriguez, H., Sood, A., Thurston, G. D., To, T., Vanker, A., and Wuebbles, D. J. (2019). Air Pollution and Noncommunicable Diseases: A Review by the Forum of International Respiratory Societies’ Environmental Committee, Part 1: The Damaging Effects of Air Pollution. *CHEST*, 155(2):409–416.

Seoul Statistics Office (2019). Seoul Vehicle Registration.

Shapiro, M. A. (2018). Transboundary air pollution in Northeast Asia: The political economy of yellow dust, particulate matter, and PM2. 5.

Shin, D.-C. (2016). *Hazardous Air Pollutants: Case Studies from Asia*. CRC Press.

Shin, H. and Bithell, M. (2019). An Agent-Based Assessment of Health Vulnerability to Long-Term Particulate Exposure in Seoul Districts. *Journal of Artificial Societies and Social Simulation*, 22(1):12.

Shou, Y., Huang, Y., Zhu, X., Liu, C., Hu, Y., and Wang, H. (2019). A review of the possible associations between ambient PM2.5 exposures and the development of Alzheimer's disease. *Ecotoxicology and Environmental Safety*, 174:344–352.

Sloan, C. D., Philipp, T. J., Bradshaw, R. K., Chronister, S., Barber, W. B., and Johnston, J. D. (2016). Applications of GPS-tracked personal and fixed-location PM2.5 continuous exposure monitoring. *Journal of the Air & Waste Management Association*, 66(1):53–65.

Smit, R., Ntziachristos, L., and Boulter, P. (2010). Validation of road vehicle and traffic emission models - A review and meta-analysis. *Atmospheric Environment*, 44(25):2943–2953.

Smith, D. M. (2012). Simulating spatial health inequalities. In *Agent-based models of geographical systems*, pages 499–510. Springer.

Smith, J. D., Barratt, B. M., Fuller, G. W., Kelly, F. J., Loxham, M., Nicolosi, E., Priestman, M., Tremper, A. H., and Green, D. C. (2020). PM2.5 on the London Underground. *Environment International*, 134:105188.

Smith, J. D., Mitsakou, C., Kitwiroon, N., Barratt, B. M., Walton, H. A., Taylor, J. G., Anderson, H. R., Kelly, F. J., and Beevers, S. D. (2016). London Hybrid Exposure Model: Improving Human Exposure Estimates to NO2 and PM2.5 in an Urban Setting. *Environmental Science and Technology*, 50(21):11760–11768.

Smith, L. and Turner, M. (2019). Building the Urban Observatory: Engineering the largest set of publicly available real-time environmental urban data in the UK. In *Geophysical Research Abstracts*, volume 21.

Son, J.-Y., Bell, M. L., and Lee, J.-T. (2010). Individual exposure to air pollution and lung function in Korea: Spatial analysis using multiple exposure approaches. *Environmental Research*, 110(8):739–749.

Song, F., Shin, J. Y., Jusino-Atresino, R., and Gao, Y. (2011). Relationships among the springtime ground–level NOx, O3 and NO3 in the vicinity of highways in the US East Coast. *Atmospheric Pollution Research*, 2(3):374–383.

Song, I. and Kim, S. Y. (2016). Estimation of Representative Area-Level Concentrations of Particulate Matter(PM10) in Seoul, Korea.

Srimuruganandam, B. and Shiva Nagendra, S. M. (2010). Analysis and interpretation of particulate matter – PM10, PM2.5 and PM1 emissions from the heterogeneous traffic near an urban roadway. *Atmospheric Pollution Research*, 1(3):184–194.

Stafoggia, M., Samoli, E., Alessandrini, E., Cadum, E., Ostro, B., Berti, G., Faustini, A., Jacquemin, B., Linares, C., and Pascal, M. (2013). Short-term associations between fine and coarse particulate matter and hospitalizations in Southern Europe: results from the MED-PARTICLES project. *Environmental health perspectives*, 121(9):1026–1033.

Stanilov, K. (2012). Space in agent-based models. In *Agent-based models of geographical systems*, pages 253–269. Springer.

Steinle, S., Reis, S., and Sabel, C. E. (2013). Quantifying human exposure to air pollution-Moving from static monitoring to spatio-temporally resolved personal exposure assessment. *Science of the Total Environment*, 443:184–193.

Steinle, S., Reis, S., Sabel, C. E., Semple, S., Twigg, M. M., Braban, C. F., Leeson, S. R., Heal, M. R., Harrison, D., Lin, C., and Wu, H. (2015). Personal exposure monitoring of PM2.5 in indoor and outdoor microenvironments. *Science of the Total Environment*, 508:383–394.

Sun, S., Cao, W., Qiu, H., Ran, J., Lin, H., Shen, C., Siu-Yin Lee, R., and Tian, L. (2020). Benefits of physical activity not affected by air pollution: a prospective cohort study. *International Journal of Epidemiology*, 49(1):142–152.

Sun, Y. and Mobasheri, A. (2017). Utilizing Crowdsourced data for studies of cycling and air pollution exposure: A case study using Strava Data. *International journal of environmental research and public health*, 14(3):274.

Taillandier, P., Vo, D.-A., Amouroux, E., and Drogoul, A. (2010). GAMA: a simulation platform that integrates geographical information data, agent-based modeling and multi-scale control. In *International Conference on Principles and Practice of Multi-Agent Systems*, pages 242–258. Springer.

Tainio, M., de Nazelle, A. J., Götschi, T., Kahlmeier, S., Rojas-Rueda, D., Nieuwenhuijsen, M. J., de Sá, T. H., Kelly, P., and Woodcock, J. (2016). Can air pollution negate the health benefits of cycling and walking? *Preventive Medicine*, 87:233–236.

TfL (2018). Speed, Emissions, and Health.

Thiele, J. C., Kurth, W., and Grimm, V. (2014). Facilitating Parameter Estimation and Sensitivity Analysis of Agent-Based Models: A Cookbook Using NetLogo and 'R'. *Journal of Artificial Societies and Social Simulation*, 17(3):11.

Timmers, V. R. J. H. and Achten, P. A. J. (2018). *Non-Exhaust PM Emissions From Battery Electric Vehicles*. Elsevier Inc.

Tonne, C., Beevers, S., Armstrong, B., Kelly, F., and Wilkinson, P. (2008). Air pollution and mortality benefits of the London Congestion Charge: spatial and socioeconomic inequalities. *Occupational and Environmental Medicine*, 65(9):620–627.

Tracy, M., Cerdá, M., and Keyes, K. M. (2018). Agent-Based Modeling in Public Health: Current Applications and Future Directions. *Annual Review of Public Health*, 39(1):77–94.

Ueberhuber, C. W. (2012). *Numerical computation 1: methods, software, and analysis*. Springer Science & Business Media.

UK Department of Health (2015). Committee on the Medical Effects of Air Pollutants. *GOV.UK Publications*, pages 1–10.

USEPA (2020). Criteria of Air Pollutants.

Van Roosbroeck, S., Jacobs, J., Janssen, N. A. H., Oldenwening, M., Hoek, G., and Brunekreef, B. (2007). Long-term personal exposure to PM2.5, soot and NOx in children attending schools located near busy roads, a validation study. *Atmospheric Environment*, 41(16):3381–3394.

Van Ryswyk, K., Wheeler, A. J., Wallace, L., Kearney, J., You, H., Kulka, R., and Xu, X. (2013). Impact of microenvironments and personal activities on personal PM2.5 exposures among asthmatic children. *Journal Of Exposure Science And Environmental Epidemiology*, 24:260.

Vreeland, H., Weber, R., Bergin, M., Greenwald, R., Golan, R., Russell, A. G., Verma, V., and Sarnat, J. A. (2017). Oxidative potential of PM2.5 during Atlanta rush hour: Measurements of in-vehicle dithiothreitol (DTT) activity. *Atmospheric Environment*, 165:169–178.

Vu, T. M., Brennan, A., Probst, C., Strong, M., Epstein, J. M., and Purshouse, R. C. (2019). Toward inverse generative social science using multi-objective genetic programming. *GECCO 2019 - Proceedings of the 2019 Genetic and Evolutionary Computation Conference*, pages 1356–1363.

Wang, S., Zhao, Y., Chen, G., Wang, F., Aunan, K., and Hao, J. (2008). Assessment of population exposure to particulate matter pollution in Chongqing, China. *Environmental Pollution*, 153(1):247–256.

Wardekker, J. A., de Jong, A., van Bree, L., Turkenburg, W. C., and van der Sluijs, J. P. (2012). Health risks of climate change: An assessment of uncertainties and its implications for adaptation policies. *Environ Health*, 11(67):1–16.

Weimer, C. W., Miller, J. O., and Hill, R. R. (2016). Agent-based modeling: An introduction and primer. In *2016 Winter Simulation Conference (WSC)*, pages 65–79. IEEE.

Weinbruch, S., Worringen, A., Ebert, M., Scheuvens, D., Kandler, K., Pfeffer, U., and Bruckmann, P. (2014). A quantitative estimation of the exhaust, abrasion and resuspension components of particulate traffic emissions using electron microscopy. *Atmospheric Environment*, 99:175–182.

WHO (2003). Health Aspects of Air Pollution with Particulate Matter , Ozone and Nitrogen Dioxide. *Report on a WHO Working Group Bonn, Germany 13–15 January 2003*, (January):98.

WHO (2012). 7 million premature deaths annually linked to air pollution.

WHO (2013). Review of evidence on health aspects of air pollution–REVIHAAP Project.

Wickham, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, 4(43):1686.

Wilensky, U. (1999). NetLogo. Evanston, IL: Center for connected learning and computer-based modeling, Northwestern University.

Wilensky, U. and Rand, W. (2015). *An introduction to agent-based modeling: modeling natural, social, and engineered complex systems with NetLogo*. MIT Press.

Wilensky, U. and Reisman, K. (2006). Thinking like a wolf, a sheep, or a firefly: Learning biology through constructing and testing computational theories—an embodied modeling approach. *Cognition and instruction*, 24(2):171–209.

Williams, M., Barrowcliffe, R., Laxen, D., and Monks, P. (2011). Review of Air Quality modelling in Defra. *A report by the Air Quality Modeling Review Steering Group*.

Wong, D. W., Yuan, L., and Perlin, S. A. (2004). Comparison of spatial interpolation methods for the estimation of air quality data. *Journal of Exposure Science & Environmental Epidemiology*, 14(5):404–415.

Wood, H. E., Marlin, N., Mudway, I. S., Bremner, S. A., Cross, L., Dundas, I., Grieve, A., Grigg, J., Jamaludin, J. B., Kelly, F. J., Lee, T., Sheikh, A., Walton, R., and Griffiths, C. J. (2015). Effects of Air Pollution and the Introduction of the London Low Emission Zone on the Prevalence of Respiratory and Allergic Symptoms in Schoolchildren in East London: A Sequential Cross-Sectional Study. *PLOS ONE*, 10(8):1–12.

Wood, S. (2019). Simplified integrated nested Laplace approximation. *Biometrika*.

Wood, S. N. (2017). *Generalized additive models: an introduction with R*. Chapman and Hall/CRC.

Wooldridge, M. (2009). *An introduction to multiagent systems*. John Wiley & Sons.

Wu, C.-Y., Mossa, J., Mao, L., and Almulla, M. (2019). Comparison of different spatial interpolation methods for historical hydrographic data of the lowermost Mississippi River. *Annals of GIS*, 25(2):133–151.

Xie, X., Semanjski, I., Gautama, S., Tsiligianni, E., Deligiannis, N., Rajan, R. T., Pasveer, F., and Philips, W. (2017). A review of urban air pollution monitoring and exposure assessment methods. *ISPRS International Journal of Geo-Information*, 6(12):1–21.

Xiong, J., Ye, C., Zhou, T., and Cheng, W. (2019). Health risk and resilience assessment with respect to the main air pollutants in sichuan. *International Journal of Environmental Research and Public Health*, 16(15).

Xu, Y., Jiang, S., Li, R., Zhang, J., Zhao, J., Abbar, S., and González, M. C. (2019). Unraveling environmental justice in ambient PM 2.5 exposure in Beijing: A big data approach. *Computers, Environment and Urban Systems*, 75(December 2018):12–21.

Yang, L., Hoffmann, P., Scheffran, J., Rühe, S., Fischereit, J., and Gasser, I. (2018). An Agent-Based Modeling Framework for Simulating Human Exposure to Environmental Stresses in Urban Areas. *Urban Science*, 2(2):36.

Young, R. L., Weinberg, J., Vieira, V., Ozonoff, A., and Webster, T. F. (2011). Generalized additive models and inflated type I error rates of smoother significance tests. *Computational statistics & data analysis*, 55(1):366–374.

Zawar-Reza, P., Kingham, S., and Pearce, J. (2005). Evaluation of a year-long dispersion modelling of PM10 using the mesoscale model TAPM for Christchurch, New Zealand. *Science of The Total Environment*, 349(1):249–259.

Zeng, W. and Church, R. L. (2009). Finding shortest paths on real road networks: the case for A*. *International Journal of Geographical Information Science*, 23(4):531–543.

Zhang, K. and Batterman, S. (2013). Air pollution and health risks due to vehicle traffic. *Science of the Total Environment*, 450-451:307–316.

Zhang, P. and Shen, T. (2015). Comparison of different spatial interpolation methods for atmospheric pollutant {PM}2.5 by using {GIS} and Spearman correlation. *Journal of Chemical and Pharmaceutical Research*, 7(12):18.

Zheng, F., Tao, R., Maier, H. R., See, L., Savic, D., Zhang, T., Chen, Q., Assumpção, T. H., Yang, P., and Heidari, B. (2018). Crowdsourcing methods for data collection in geophysics: State of the art, issues, and future directions. *Reviews of Geophysics*, 56(4):698–740.

Zheng, H., Son, Y.-J., Chiu, Y.-C., Head, L., Feng, Y., Xi, H., Kim, S., and Hickman, M. (2013). A Primer for Agent-Based Simulation and Modeling in Transportation Applications. Technical report.

Żak, M., Melaniuk-Wolny, E., and Widziewicz, K. (2017). The exposure of pedestrians, drivers and road transport passengers to nitrogen dioxide. *Atmospheric Pollution Research*, 8(4):781–790.

Appendix A

A Brief Exploration of Dispersion Models

A.1 Model Introduction

A.1.1 Air Pollution Dispersion Modelling System (ADMS)

ADMS is a steady-state, three dimensional, quasi-Gaussian, regulatory model from the UK which aims to measure the dispersion of buoyant particles and gases in large urban areas, cities, and towns (Carruthers et al., 1994; Di Sabatino et al., 2008; Holmes and Morawska, 2006).

The Cambridge Environmental Research Consultants Corp. (CERC) initiated this model in 1994, investigating point source plumes from industrial sites (ADMS 5). Fortunately, CERC has expanded models that analyse emissions from road (ADMS-roads), urban area (ADMS-Urban), and airport (ADMS-Airport), and has worked with different countries across the world: France, Malta, China, Hong Kong, Singapore, South Africa and USA. Above all, ADMS-Urban is widely used today, with its last update in November 2016.

ADMS is a quasi-Gaussian plume model because the pioneers of ADMS developed this model by improving the "steady state air quality" assumptions run by earlier models of Pasquill (1961). Carruthers et al. (1994) argued that steady state conditions of air oversimplified results by classifying ground level meteorological conditions to the entire boundary layer, without considering building downwashes or any barrier

conditions that would happen in the real world. As a solution, the developers applied Monin-Obukhov equation to calculate the effects of the buoyancy on turbulent flows within the boundary layer[1].

ADMS covers atmospheric dispersion in point, line, area, and volume sources in an hourly basis. The point source of the model considers plume dispersion and building downwash –up to 25 buildings per model–, and dry/wet depositions which has a relatively basic module (Di Sabatino et al., 2008). The line source component is use to model road level dispersion, applying the street canyon convolution equation. The area and volume dispersion are modelled by an advance function of ADMS-Urban package. The model has an ability to produce a maximum number of 300 emission sources, with limits of 30 line sources, 30 area sources, and 30 volume sources in a model. ADMS-Urban produces 10 pollutant outputs in a single iteration, to a grid of receptor points, along with descriptions of basic statistics. Visualisation is available on their local software ADMS Mapper, or on ArcGIS –where ADMS calculation extension is installed.

A.1.2 Community Multiscale Air Quality (CMAQ): Eulerian Model

CMAQ is a three-dimensional air quality model to estimate concentration of multiple pollutants at different geographic scales (Beevers et al., 2013; Bravo et al., 2012; Byun and Schere, 2006; Hoek et al., 2008). CMAQ, pronounced as see-mak, was conceived with the intention of building a 'one atmosphere' and a high quality model which can help to observe air quality over a wide length of time series (Williams et al., 2011). Also, the word 'community' is to offer different options for the user's intentions, e.g. emission, meteorological conditions, or to simply mean an open-source platform. US-EPA established CMAQ as a regulatory model to suggest policy evaluation and to assess control of air quality standards (Bravo et al., 2012).

As mentioned in the previous section, CMAQ uses an Eulerian approach that decomposes space into 3D Euclidean grid cells, from 1km by 1km up to 36km by 36km, to calculate the process of emission diffusion, chemical reaction, and dissipation of pollutants within a given boundary (Bravo et al., 2012; Byun and Schere, 2006). The model has fast calculation speed to produce particles, ozone, and acid depositions, and

[1]Boundary layer: a vertical layer affected by the solar heating system of the earth surface

the user can perform simulation in either a long-term (e.g. year to multiple years) or a short-term period (e.g. days to months). However, the computational speed is much faster when it runs on a HPC, such as Darwin cluster in the University of Cambridge.

CMAQ incorporates three kinds of models: 1) meteorological models to simulate atmospheric and weather phenomena; 2) emission models to simulate artificial and natural sources to the atmosphere; and 3) an air chemistry-transport model to simulate the dissipation rate of pollutants under various conditions.

Previous studies have also discussed the relationship between health impacts and ambient air pollution, particularly ozone (Bell, 2006; Knowlton et al., 2004), both ozone and $PM_{2.5}$ (Bravo et al., 2012; Fann et al., 2012), or both nitrogen dioxides and $PM_{2.5}$ (Beevers et al., 2013; Jung et al., 2016; Smith et al., 2016). Reviewed research was conducted in USA (mainly northeast), UK, and Korea, but more studies have been carried out in almost 50 countries.

Knowlton et al. (2004) investigated the summer ozone concentrations and its relation to mortality, with CMAQ's dispersion model SMOKE and weather forecast model MM5. The authors concluded that the rise of ozone projected an increase of mortality rate by around 4.5% for New York by 2050 if USA continues to generate equal amounts of airborne emissions from power plants, or continues to over-consume air conditioners or dispose waste at landfill sites. Similar work was conducted by Fann et al. (2012), where CMAQ was used to estimate $PM_{2.5}$ and 8-hour-ozone levels with a 12km grid cell, and examine the relationship between statistics of premature death and life expectancy in the United States. The result predicted 130,000 deaths to $PM_{2.5}$, and 4,700 to ozone, and over 1.1 million of elderly people are more likely to have their life expectancy reduced. Also, Fann et al. (2012) gave an interesting rank of mortality rate of cities that were affected by ozone and $PM_{2.5}$ in the year of 2005. However, these results might have over-predicted risks in terms of interpreting pollution concentration associated with mortality figures. Hence, more consideration before examining the link between mortality and air quality standards is needed.

In fact, some epidemiology studies used CMAQ without putting health measures forward (Bell, 2006; Bravo et al., 2012). These papers have used CMAQ modelling in order to fill in area gaps for estimating exposure levels where monitoring stations do not exist. These studies were possible because CMAQ improved the spatial resolution down to 1km, which can verify the data with more than 2000 stations. Hourly-based time series also gave a detailed pollution pattern. Bell (2006) estimated exposure levels

with multiple approaches, including area-weighted average pollution using CMAQ, nearest monitor, and spatial interpolation (i.e. IDW, Kriging). They concluded that previous methods gave a poor value at places where monitoring stations were sparse, and suggested the need to thoroughly examine modelling methods before measuring personal exposure. Bravo et al. (2012) also simulated pollution dispersion by aggregating the values from 1km grid cells, and compared the modelled values with populations where monitoring stations are close or not. Results showed people who reside within 50km of the station have higher incomes, higher academic degrees, higher employment rates, and a lower poverty index. In short, CMAQ has strongly contributed to pollution and health studies with respect to technical improvement. However, area-weighted average values in 1km grid cell are too coarse to detect the dispersion concentration in an intra-urban scale, and suggests where new measuring stations should be installed, either in a regular distance or population-derived (Song and Kim, 2016).

As more pollutants originate from vehicles, more recent research (Beevers et al., 2013; Halonen et al., 2016; Smith et al., 2016) have highlighted a new trend: a hybrid model of CMAQ and a road level model. The CMAQ-urban model, established by KCLurban epidemiology department, coupled the roads model (ADMS 2.3) with weather forecasting model (WRF 3.1), and CMAQ 4.7.1 in order to predict pollution levels for London. To date, this model provides the highest resolution of dispersion prediction to NO-NO_2-O_3, PM_{10}, and $PM_{2.5}$, across a 20m by 20m grid (see details on Chapter 2 Section 2.4.2).

A.1.3 The Air Pollution Model (TAPM): Eulerian Model

TAPM is a three-dimensional, prognostic, numerical, PC-based, meso-scale dispersion model that simulates meteorological and pollution concentration in areas where the meteorological gridded data is coarse (David and Don, 2012; Hurley, 2002; Zawar-Reza et al., 2005) . Developed by the Commonwealth Scientific and Industrial Research Organisation (CSIRO) Atmospheric research team, TAPM has mostly been used in Australia and New Zealand where many cities are located near the coast, affected by coastal temperature, also with complex terrains.

The model runs on a GUI platform with a gridded map plotted in the middle of the interface (Figure 11). Geographical extent decomposes to 1500km x 1500km lattice,

with each grid of 1.5km2. Input data include terrain height, land cover, global soil texture, sea surface temperature, wind speed and direction, temperature, humidity and so on. More information is provided in Bluett et al. (2004); Hurley (2008).

There were several projects and publications based on TAPM. Zawar-Reza et al. (2005) examined annual concentration of PM10 for 1999 in Christchurch, New Zealand. Although it was a pilot version, the pre-processing work for traffic, industrial emissions and wind roses were thoroughly examined. The modelled results showed a great fit to daily PM_{10} concentration during most of the days in January, but less fit in August because the model did not consider emissions from home heating or natural emissions. Overall, the model predicted the daily average of PM_{10} concentration well in Christchurch, as the evidence of the Index of Agreement (IOA) scores were over 0.6 (Values close to 1 equals high accuracy).

A.2 A Comparison of Dispersion Models

A.2.1 ADMS

Pros. ADMS-Road is the widely used package for road dispersion models. The software is based on a Windows platform, which allows a user-friendly interface of input data and results with only a few clicks. ADMS give a better understanding of the association between emission sources and pollution concentration with separate packages. For example, ADMS-Roads to estimate NO_2 concentration in roads.

Cons. The initial critique, as mentioned earlier, is that it has less capability of modelling the contribution of distance sources (Williams et al., 2011). Several studies have already modelled the regional effect of chemical transport with meso-scale models including CMAQ. In terms of modelling speed, it is too slow to model a city with an ADMS-Urban package probably due to the model's single core design, or an unsorted algorithm when blending two different packages . Finally, ADMS is for commercial use only, which means a certain amount of money needs to be paid for an annual license at least £3,000 for annual license plus tutorial classes.

A.2.2 CMAQ

Pros. CMAQ has great advantage as an open source model. This indicates that the model is transparent and can be customised with other emission or meteorological models. Another remarkable feature is its flexibility of 'nesting' multiple models across urban to regional scale. The recent version (CMAQ 5.2) allows users to freely utilise its spatial scale at $1km^2$, $2km^2$, $5km^2$, and $10km^2$, and furthermore, can explicitly adapt to a roadside module (e.g. ADMS-Roads).

Cons. Despite these advantages, this model is relatively difficult to use. The default mode includes emission inventory, weather forecast, and chemical transport modules in which each model itself has a complicated algorithm. A continuous concern is its computation intensity. Fortunately, the upside is that Cambridge has the HPC cluster to help accelerating considerable time and resources.

A.2.3 TAPM

Pros. The advantage of TAPM is that the model speed is fast. With a Eulerian approach, the model automatically constructs a three-dimensional Cartesian grid for point source pollutants, and calculates dispersion. A Lagrangian option is also available for gradual plume rise and near-source effects (Hurley, 2008). Another feature of TAPM is that it has a GUI to help users for different level of proficiency in programming to approach it with ease. In fact, every sequence is very simple that the data are simply imported and exported with only a few clicks, making the users trace it without difficulty.

Cons. The model only activates within the spatial boundary, because it does not include the curvature of the earth. Moreover, a $1km^2$ grid resolution also restricts examination for perturbation effects at a high level of spatial detail. Secondly, the model cannot simulate steep terrain because the model does not take into consideration discontinuity of height. Furthermore, it has less accuracy of measuring extreme weather events and atmospheric circulations due to the incompressibility of the model (Hurley, 2008). Finally, and unfortunately, this model is no longer available.

Table A.1 Summary of Dispersion Models

Type	ADMS	CMAQ	TAPM
Full Name	Air Dispersion Modelling System	Community Multiscale Air Quality	The Air Pollution Model
Institution	CERC, UK	EPA, USA	CSIRO, Australia
Module	Quasi-Gaussian	Eulerian	Lagrangian / Eulerian
Scale	Street canyon	National / Continental	Meso
Resolution	20m x 20m	1km / 3km / 12km / 36km	Grid: 0.3 km to 30 km
Measured Pollutants	PM_{10}, $PM_{2.5}$, NO_x, SO_x, VOCs, Benzene, O_3	PM_{10}, $PM_{2.5}$, NO_x, SO_x, O_3	PM_{10}, $PM_{2.5}$, TSP, CO, SO_x, NO_x, VOCs, O_3
System	Windows	Linux	Windows
CPU	Single Core	Multiple Core (Parallel)	Single Core

Appendix B

NetLogo Codes

B.1 Chapter 4: A Case Study of Gangnam

```
extensions [gis csv table]
globals [gu road land lc districtPop districtadminCode districtEdu \%riskpop date
↪  where number-dead poll_scenario ]
breed [dong-labels dong-label]
breed[people person]
patches-own [is-research-area? is-built-area? name dong-code land10 hospital]
people-own  [health age edu districtName district-code
             homeName homePatch destinationName destinationPatch]

#;;---------------------------------
to setup
 clear-all
 reset-ticks
 set-gis-data
 add-labels
 add-admin
 add-land
 add-census
 add-pollution
 set-dictionaries
 set-people
 set-destination
end
#;;---------------------------------
to go
```

```
  calc-pm10
  landprice-change
  move-people
  ask people [
  inhalation
  ]
  go-hospital
  ask people [
    adaptive-cap
    sensitivity
    ]

  gu-plot
  dong-plot
  age-plot
  edu-plot
  pm10-plot
  update-plots
  tick
  if (ticks = 8764) [stop]
  set date item 0 table:get ts_kalman (ticks + 1)
  set where item 2 table:get ts_kalman (ticks + 1)
  set \%riskpop    (count people with [color = red and destinationName != "others"]
  ↪  / count people with [destinationName != "others"]) * 100
end


#;;--------------------------------
to set-gis-data
  ask patches [set pcolor white]
  gis:load-coordinate-system (word "boundary/boundary_shape/Gangnam.prj")
  set gu   gis:load-dataset "boundary/boundary_shape/Gangnam.shp"
  set lc   gis:load-dataset "landcover/Gangnam.shp"
  set road gis:load-dataset "roads/roads_shape/gangnam.shp"
  set land gis:load-dataset "LandPrice/LandPrice_Gu_Shape/Landprice_Gangnam.shp"
  gis:set-world-envelope (gis:envelope-union-of gis:envelope-of gu)
  ask patches gis:intersecting gu [set is-research-area? true]   ask patches
  ↪  gis:intersecting lc [set is-built-area? true]
#;;--------------------------------
  #;; Draw district
foreach gis:feature-list-of gu [ gu-feature ->
```

```
    gis:set-drawing-color scale-color green (gis:property-value gu-feature
    ↪ "ADM_CD") 1123051 1123080
    gis:fill gu-feature 0
  ]
  gis:set-drawing-color [  64  64  64]    gis:draw gu 1
  gis:set-drawing-color [255 0 0 100]    gis:draw road 1
end

#;;--------------------------------
to add-labels
  foreach gis:feature-list-of gu [vector-feature ->
  let centroid gis:location-of gis:centroid-of vector-feature
       if not empty? centroid
      [ create-dong-labels 1
        [ set xcor item 0 centroid
          set ycor item 1 centroid
          set size 0
          set label-color blue
          set label gis:property-value vector-feature "ADM_NM"
        ]]]
end
#;;--------------------------------
to add-admin
  gis:set-drawing-color blue
  foreach gis:feature-list-of gu [vector-feature ->
    ask patches[ if gis:intersects? vector-feature self [set name
    ↪ gis:property-value vector-feature "ADM_NM"
                                  set dong-code gis:property-value vector-feature
                                  ↪ "ADM_CD"]
 ]]
end
#;;--------------------------------
to add-land
  gis:set-drawing-color blue
  foreach gis:feature-list-of land [vector-feature ->
    ask patches[ if gis:intersects? vector-feature self [set land10
    ↪ gis:property-value vector-feature "pr10"
                                  ]
 ]]
end
#;;--------------------------------
to add-census
```

```
  let rawCode csv:from-file "Census/census2010_age_5per.csv"
  let adCode table:make
  foreach rawCode [ code ->
         if item 1 code = "Gangnam"
        [table:put adCode item 0 code list (item 1 code)(item 2 code) ]
        ]
end

to add-pollution
#; Import daily pollution
  let p0 csv:from-file "Pollution/St111261_Gangnam.csv"
  let p1 remove-item 0 p0
  let rep 0
  set ts_kalman  table:make

  foreach p1 [ poll ->
  if item 2 poll = "pm10"
    [ let counter item 0 poll
      let the-rest remove-item 0 poll
      table:put ts_kalman counter the-rest
    ]
  ]
set rep rep + 1

  ask patches with [is-research-area? = true] [
    let homeID item (3 + random 13) table:get ts_kalman 1
    ifelse homeID > 0
    [set ts__kal  homeID][set ts__kal max table:get ts_kalman 1]
  ]


#;;Scenarios
  let quarter csv:from-file "Scenarios/St111261_Gangnam.csv"
  let q1 remove-item 0 quarter
  let looop 0
  set poll_scenario table:make

   foreach q1 [p ->
    if item 1 p = "pm10"
    [ let counter item 0 p
      let the-rest remove-item 0 p
      table:put poll_scenario counter the-rest]
```

```
  ]
  set looop looop + 1

end

#;;;
```

```
     let high    lput item 5 code middle
     let college lput item 6 code high
     let univ    lput item 7 code college
     let master  lput item 8 code univ
     let phd     lput item 9 code master
     table:put districtEdu item 1 code phd
    ]
  ]
end

#;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;
to set-people
  foreach table:keys districtpop [ dist ->
    let ageGroupID 0
    foreach table:get districtpop dist [ number ->
      create-people number [
        setupAgeGroup agegroupID
        set districtName dist
        set district-code table:get districtadminCode dist
        set shape "person"
        set heading random 360
        set homeName dist
        set homePatch one-of patches with [dong-code = [district-code] of myself
        ↪  and is-built-area? = true]
        move-to homePatch
        set destinationName "unidentified"
        set destinationPatch "unidentified"
        set health 300
      ]
      set ageGroupID AgeGroupID + 1
      ]
  ]
end

to setupAgeGroup [ID]
if ID = 0  [set size 1 set age  5 + random 5 set color orange
            ifelse random-float 1 <= .18 [set edu 0][set edu 1]]
if ID = 1  [set size 1 set age 10 + random 5 set color orange + 1
            ifelse random-float 1 <= .49 [set edu 1][set edu 2]]
if ID = 2  [set size 1 set age 15 + random 5 set color orange + 2
            let r random-float 1
            if (0    < r and r <= 0.14)[set edu 2]
```

```
            if (0.14 < r and r <= 0.84)[set edu 3]
            if (0.84 < r and r <= 0.88)[set edu 4]
            if (0.88 < r )[set edu 5]]
if ID = 3  [set size 1 set age 20 + random 5 set color turquoise
            let r random-float 1
            if (0    < r and r <= 0.11)[set edu 3]
            if (0.11 < r and r <= 0.31)[set edu 4]
            if (0.31 < r and r <= 0.98)[set edu 5]
            if (0.98 < r )[set edu 6]]
if ID = 4  [set size 1 set age 25 + random 5 set color turquoise
            let r random-float 1
            if (0    < r and r <= 0.11)[set edu 3]
            if (0.11 < r and r <= 0.28)[set edu 4]
            if (0.28 < r and r <= 0.88)[set edu 5]
            if (0.88 < r and r <= 0.98)[set edu 6]
            if (0.98 < r )[set edu 7]]
if ID = 5  [set size 1 set age 30 + random 5 set color turquoise
            let r random-float 1
            if (0    < r and r <= 0.12)[set edu 3]
            if (0.12 < r and r <= 0.27)[set edu 4]
            if (0.27 < r and r <= 0.82)[set edu 5]
            if (0.82 < r and r <= 0.96)[set edu 6]
            if (0.96 < r )[set edu 7]]
if ID = 6  [set size 1 set age 35 + random 5 set color turquoise
            let r random-float 1
            if (0    < r and r <= 0.15)[set edu 3]
            if (0.15 < r and r <= 0.28)[set edu 4]
            if (0.28 < r and r <= 0.77)[set edu 5]
            if (0.77 < r and r <= 0.94)[set edu 6]
            if (0.94 < r )[set edu 7]]
if ID = 7  [set size 1 set age 40 + random 5 set color brown
            let r random-float 1
            if (0    < r and r <= 0.01)[set edu 2]
            if (0.01 < r and r <= 0.17)[set edu 3]
            if (0.17 < r and r <= 0.27)[set edu 4]
            if (0.27 < r and r <= 0.77)[set edu 5]
            if (0.77 < r and r <= 0.93)[set edu 6]
            if (0.93 < r )[set edu 7]]
if ID = 8  [set size 1 set age 45 + random 5 set color brown
            let r random-float 1
            if (0    < r and r <= 0.01)[set edu 1]
            if (0.01 < r and r <= 0.03)[set edu 2]
```

```
            if (0.03 < r and r <= 0.22)[set edu 3]
            if (0.22 < r and r <= 0.31)[set edu 4]
            if (0.31 < r and r <= 0.79)[set edu 5]
            if (0.79 < r and r <= 0.93)[set edu 6]
            if (0.93 < r )[set edu 7]]
if ID = 9  [set size 1 set age 50 + random 5 set color brown
            let r random-float 1
            if (0    < r and r <= 0.02)[set edu 1]
            if (0.02 < r and r <= 0.06)[set edu 2]
            if (0.06 < r and r <= 0.32)[set edu 3]
            if (0.32 < r and r <= 0.40)[set edu 4]
            if (0.40 < r and r <= 0.81)[set edu 5]
            if (0.81 < r and r <= 0.92)[set edu 6]
            if (0.92 < r )[set edu 7]]
if ID = 10 [set size 1 set age 55 + random 5 set color brown
            let r random-float 1
            if (0    < r and r <= 0.04)[set edu 1]
            if (0.04 < r and r <= 0.11)[set edu 2]
            if (0.11 < r and r <= 0.39)[set edu 3]
            if (0.39 < r and r <= 0.46)[set edu 4]
            if (0.46 < r and r <= 0.85)[set edu 5]
            if (0.85 < r and r <= 0.92)[set edu 6]
            if (0.92 < r )[set edu 7]]
if ID = 11 [set size 1 set age 60 + random 5 set color violet
            let r random-float 1
            if (0    < r and r <= 0.01)[set edu 0]
            if (0.01 < r and r <= 0.08)[set edu 1]
            if (0.08 < r and r <= 0.16)[set edu 2]
            if (0.16 < r and r <= 0.43)[set edu 3]
            if (0.43 < r and r <= 0.49)[set edu 4]
            if (0.49 < r and r <= 0.89)[set edu 5]
            if (0.89 < r and r <= 0.97)[set edu 6]
            if (0.97 < r )[set edu 7]]
if ID = 12 [set size 1 set age 65 + random 5 set color violet
            let r random-float 1
            if (0    < r and r <= 0.02)[set edu 0]
            if (0.02 < r and r <= 0.12)[set edu 1]
            if (0.12 < r and r <= 0.21)[set edu 2]
            if (0.21 < r and r <= 0.46)[set edu 3]
            if (0.46 < r and r <= 0.52)[set edu 4]
            if (0.52 < r and r <= 0.89)[set edu 5]
            if (0.89 < r and r <= 0.96)[set edu 6]
```

```
            if (0.96 < r )[set edu 7]]
if ID = 13 [set size 1 set age 70 + random 5 set color violet
            let r random-float 1
            if (0    < r and r <= 0.05)[set edu 0]
            if (0.05 < r and r <= 0.21)[set edu 1]
            if (0.21 < r and r <= 0.31)[set edu 2]
            if (0.31 < r and r <= 0.54)[set edu 3]
            if (0.54 < r and r <= 0.59)[set edu 4]
            if (0.59 < r and r <= 0.90)[set edu 5]
            if (0.90 < r and r <= 0.96)[set edu 6]
            if (0.96 < r )[set edu 7]]
if ID = 14 [set size 1 set age 75 + random 5 set color violet
            let r random-float 1
            if (0    < r and r <= 0.10)[set edu 0]
            if (0.10 < r and r <= 0.34)[set edu 1]
            if (0.34 < r and r <= 0.45)[set edu 2]
            if (0.45 < r and r <= 0.68)[set edu 3]
            if (0.68 < r and r <= 0.72)[set edu 4]
            if (0.72 < r and r <= 0.94)[set edu 5]
            if (0.94 < r and r <= 0.98)[set edu 6]
            if (0.98 < r )[set edu 7]]
if ID = 15 [set size 1 set age 80 + random 5 set color pink
            let r random-float 1
            if (0    < r and r <= 0.17)[set edu 0]
            if (0.17 < r and r <= 0.48)[set edu 1]
            if (0.48 < r and r <= 0.60)[set edu 2]
            if (0.60 < r and r <= 0.79)[set edu 3]
            if (0.79 < r and r <= 0.84)[set edu 4]
            if (0.84 < r and r <= 0.97)[set edu 5]
            if (0.97 < r and r <= 0.99)[set edu 6]
            if (0.99 < r )[set edu 7]]
if ID = 16 [set size 1 set age 85 + random 15 set color pink
            let r random-float 1
            if (0    < r and r <= 0.29)[set edu 0]
            if (0.29 < r and r <= 0.62)[set edu 1]
            if (0.62 < r and r <= 0.71)[set edu 2]
            if (0.71 < r and r <= 0.86)[set edu 3]
            if (0.86 < r and r <= 0.89)[set edu 4]
            if (0.89 < r and r <= 0.98)[set edu 5]
            if (0.98 < r and r <= 0.99)[set edu 6]
            if (0.99 < r )[set edu 7]]
end
```

```
#;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;
to set-destination   #;; Decomposing matrix
  let gncsv csv:from-file "ODmatrix/St111261_Gangnam.csv"
  let rawheader item 0 gncsv
  let destinationNames remove-item 0 rawheader
  let gnMat remove-item 0 gncsv

  let loopnum 1
  let gnMatrix table:make #;; This is a matrix where each origin has its name as a
  ↪  "key"
  foreach gnMat [ origin-chart ->
    let numberMat remove-item 0 origin-chart    #;; fraction has to be btw 0-1,
    let fraction map [ i -> i / 100 ] numberMat #;; but the original file is btw
    ↪  1-100
    table:put gnMatrix item 0 origin-chart fraction
                ]
  set loopnum loopnum + 1

  foreach table:keys gnMatrix [ originName ->
    let matrix-loop 0
    let Num count people with [homeName = originName and (age >= 15 and age < 65)]
    let totalUsed 0
    let number 0

        foreach table:get gnMatrix originName
       [ percent ->
          let newDestination item matrix-loop destinationNames #;; Let agents of 22
          ↪  origins choose their destinations
          ifelse (newDestination != "others") [set number precision (percent * Num)
          ↪  0 set totalUsed totalUsed + number]
              [set number Num - totalUsed ]
                  #;; if agents move within district, then count agents by rounding
                  ↪  the values of population x
      #;; "fraction of region A", population x "fraction of region B"...
      #;; if agents move outside district, then count the remainder of the
      ↪  population not used for inbound population
         let peopleRemaining (people with [homeName = originName and
         ↪  destinationName = "unidentified"
               and (age >= 15 and age < 65)])
```

```
        if count peopleRemaining > 0 and count peopleRemaining <= number [ set
        ↪  number count peopleRemaining ]
             if number < 0 [ set number 0]

        ask n-of number peopleRemaining [
               set destinationName newDestination #;; assign destination name
               set destinationPatch one-of patches with [name = newDestination and
               ↪  is-built-area? = true]
      ]
    set matrix-loop matrix-loop + 1
  ]
  type totalused type " " type Num type " " print originName #;; print inbound
  ↪  agents out of the total population (age 15-64)
  ]


#;; Send agents selected as "others" to the NE corner
 ask people [ if destinationName = "others"
             [ set destinationPatch patch max-pxcor max-pycor]
             ]

 ask people with [destinationpatch = "unidentified" and age < 15]
                 [set destinationName homeName
                  set destinationPatch one-of patches in-radius 3] #;; Under 15
 ask people with [destinationpatch = "unidentified" and age >= 65]
                 [set destinationName homeName
                  set destinationPatch one-of patches in-radius 1] #;; Over 65

 output-print "People without destinations(nobody)"
 let wordloop 0
 foreach destinationNames [ dn ->
    output-print word (word(word (word dn ": ") count people with
    [homename = dn and destinationPatch = nobody] ) " out of " ) count people with
    [homename = dn ]
       ]
 set wordloop wordloop + 1

end


to set-at-hospital
  ask patch min-pxcor min-pycor [set pcolor grey + 1]
```

```
end


;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;
to go-hospital
  ask people [if (health <= 20) [move-to patch min-pxcor min-pycor fd 1]]
end

to move-people
  ifelse ticks mod 2 = 0 [move-out][come-home]
end

to move-out
  ask people [if patch-here != destinationPatch [ move-to destinationPatch fd 1]
  ]
end

to come-home
  ask people [
    if patch-here != homePatch [move-to homePatch fd 1]
  ]
end


;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;
to adaptive-cap
  if (health < AC) and ([land10] of patch-here  < 920459)  [set health health +
  ↪  0.01]
  if (health < AC) and ([land10] of patch-here >= 920459)   and ([land10] of
  ↪  patch-here < 1550940) [set health health + 0.01]
  if (health < AC) and ([land10] of patch-here >= 1550940)  and ([land10] of
  ↪  patch-here < 2091133) [set health health + 0.02]
  if (health < AC) and ([land10] of patch-here >= 2091133)  and ([land10] of
  ↪  patch-here < 2637070) [set health health + 0.04]
  if (health < AC) and ([land10] of patch-here >= 2637070)  and ([land10] of
  ↪  patch-here < 3273274) [set health health + 0.05]
  if (health < AC) and ([land10] of patch-here >= 3273274)  and ([land10] of
  ↪  patch-here < 4140183) [set health health + 0.05]
  if (health < AC) and ([land10] of patch-here >= 4140183)  and ([land10] of
  ↪  patch-here < 5443608) [set health health + 0.06]
  if (health < AC) and ([land10] of patch-here >= 5443608)  and ([land10] of
  ↪  patch-here < 8361806) [set health health + 0.07]
```

```
  if (health < AC) and ([land10] of patch-here >= 8361806)  and ([land10] of
  ↪  patch-here < 11545447) [set health health + 0.12]
  if (health < AC) and ([land10] of patch-here >= 11545447) and ([land10] of
  ↪  patch-here < 20261596) [set health health + 0.15]

end
#;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;
to inhalation
  ifelse ([road] of patch-here = true)[road-effect] [non-road-effect]
end

#;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;
to sensitivity
  if (ts__kal >= PM10-parameters) and (health < 300 or edu < 1 + random 3) and (age
  ↪  < 10)
     [set health (health - random-float (health-loss * health-loss2 * 1.0529) *
     ↪  (310 - health))]
  if (ts__kal >= PM10-parameters) and (health < 300 or edu < 1 + random 3) and (age
  ↪  >= 30 and age < 35)
     [set health (health - random-float (health-loss * health-loss2 * 0.233) * (310
     ↪  - health))]
  if (ts__kal >= PM10-parameters) and (health < 300 or edu < 1 + random 3) and (age
  ↪  >= 35 and age < 40)
     [set health (health - random-float (health-loss * health-loss2 * 0.2253) *
     ↪  (310 - health))]
  if (ts__kal >= PM10-parameters) and (health < 300 or edu < 1 + random 3) and (age
  ↪  >= 40 and age < 45)
     [set health (health - random-float (health-loss * health-loss2 * 0.2377) *
     ↪  (310 - health))]
  if (ts__kal >= PM10-parameters) and (health < 300 or edu < 1 + random 3) and (age
  ↪  >= 45 and age < 50)
     [set health (health - random-float (health-loss * health-loss2 * 0.7318) *
     ↪  (310 - health))]
  if (ts__kal >= PM10-parameters) and (health < 300 or edu < 1 + random 3) and (age
  ↪  >= 50 and age < 55)
     [set health (health - random-float (health-loss * health-loss2 * 0.5472) *
     ↪  (310 - health))]
  if (ts__kal >= PM10-parameters) and (health < 300 or edu < 1 + random 3) and (age
  ↪  >= 55 and age < 60)
     [set health (health - random-float (health-loss * health-loss2 * 1.732) * (310
     ↪  - health))]
```

```
  if (ts__kal >= PM10-parameters) and (health < 300 or edu < 1 + random 3) and (age
  ↪  >= 60 and age < 65)
     [set health (health - random-float (health-loss * health-loss2 * 0.723) * (310
     ↪  - health))]
  if (ts__kal >= PM10-parameters) and (health < 300 or edu < 1 + random 3) and (age
  ↪  >= 65 and age < 70)
     [set health (health - random-float (health-loss * health-loss2 * 1.03) * (310
     ↪  - health))]
  if (ts__kal >= PM10-parameters) and (health < 300 or edu < 1 + random 3) and (age
  ↪  >= 70 and age < 75)
     [set health (health - random-float (health-loss * health-loss2 * 1.381) * (310
     ↪  - health))]
  if (ts__kal >= PM10-parameters) and (health < 300 or edu < 1 + random 3) and (age
  ↪  >= 75)
     [set health (health - random-float (health-loss * health-loss2 * 1.602) * (310
     ↪  - health))]

  if (health < 200 and health >= 100) [set color violet]
  if (health < 100) [set color red]
end

#;;;
```

```
  [ask patches with [is-research-area? = true]
    [
     if ticks > 0 and ticks <= 4382 [set-BAU]
     if ticks > 4382 [set-INC&DEC]
    ]
  ]

  if (Scenario = "DEC")
  [ask patches with [is-research-area? = true]
    [
     if ticks > 0 and ticks <= 4382 [set-BAU]
     if ticks > 4382 [set-INC&DEC]
    ]
  ]
end

to set-BAU
  let homeID item (3 + random 13) table:get ts_kalman ticks
  let workID item (3 + random 11) table:get ts_kalman ticks + 1
  if (ticks + 1) mod 2 = 0 [set ts__kal  homeID]
  if ticks mod 2 = 0 [set ts__kal  workID]

end

to set-INC&DEC
  let homeID item (3 + random 13) table:get ts_kalman ticks
  let workID item (3 + random 11) table:get ts_kalman ticks + 1
  let \%3inc 5
  if scenario-percent = "inc-sce" [set \%3inc \%3inc]
  if scenario-percent = "dec-sce" [set \%3inc \%3inc + 1]

if (ticks > 4382 and ticks <= 4562 and ticks mod 2 = 0)[set ts__kal workID + (item
↪  \%3inc table:get poll_scenario 1)]
if (ticks > 4562 and ticks <= 4744 and ticks mod 2 = 0)[set ts__kal workID + (item
↪  \%3inc table:get poll_scenario 2)]
if (ticks > 4744 and ticks <= 4850 and ticks mod 2 = 0)[set ts__kal workID + (item
↪  \%3inc table:get poll_scenario 3)]
if (ticks > 4850 and ticks <= 5112 and ticks mod 2 = 0)[set ts__kal workID + (item
↪  \%3inc table:get poll_scenario 4)]
if (ticks > 5112 and ticks <= 5292 and ticks mod 2 = 0)[set ts__kal workID + (item
↪  \%3inc table:get poll_scenario 5)]
```

```
if (ticks > 5292 and ticks <= 5474 and ticks mod 2 = 0)[set ts__kal workID + (item
↪  \%3inc table:get poll_scenario 6)]
if (ticks > 5474 and ticks <= 5658 and ticks mod 2 = 0)[set ts__kal workID + (item
↪  \%3inc table:get poll_scenario 7)]
if (ticks > 5658 and ticks <= 5842 and ticks mod 2 = 0)[set ts__kal workID + (item
↪  \%3inc table:get poll_scenario 8)]
if (ticks > 5842 and ticks <= 6024 and ticks mod 2 = 0)[set ts__kal workID + (item
↪  \%3inc table:get poll_scenario 9)]
if (ticks > 6024 and ticks <= 6206 and ticks mod 2 = 0)[set ts__kal workID + (item
↪  \%3inc table:get poll_scenario 10)]
if (ticks > 6206 and ticks <= 6390 and ticks mod 2 = 0)[set ts__kal workID + (item
↪  \%3inc table:get poll_scenario 11)]
if (ticks > 6390 and ticks <= 6574 and ticks mod 2 = 0)[set ts__kal workID + (item
↪  \%3inc table:get poll_scenario 12)]
if (ticks > 6574 and ticks <= 6754 and ticks mod 2 = 0)[set ts__kal workID + (item
↪  \%3inc table:get poll_scenario 13)]
if (ticks > 6754 and ticks <= 6936 and ticks mod 2 = 0)[set ts__kal workID + (item
↪  \%3inc table:get poll_scenario 14)]
if (ticks > 6936 and ticks <= 7120 and ticks mod 2 = 0)[set ts__kal workID + (item
↪  \%3inc table:get poll_scenario 15)]
if (ticks > 7120 and ticks <= 7304 and ticks mod 2 = 0)[set ts__kal workID + (item
↪  \%3inc table:get poll_scenario 16)]
if (ticks > 7304 and ticks <= 7484 and ticks mod 2 = 0)[set ts__kal workID + (item
↪  \%3inc table:get poll_scenario 17)]
if (ticks > 7484 and ticks <= 7666 and ticks mod 2 = 0)[set ts__kal workID + (item
↪  \%3inc table:get poll_scenario 18)]
if (ticks > 7666 and ticks <= 7850 and ticks mod 2 = 0)[set ts__kal workID + (item
↪  \%3inc table:get poll_scenario 19)]
if (ticks > 7850 and ticks <= 8034 and ticks mod 2 = 0)[set ts__kal workID + (item
↪  \%3inc table:get poll_scenario 20)]
if (ticks > 8034 and ticks <= 8214 and ticks mod 2 = 0)[set ts__kal workID + (item
↪  \%3inc table:get poll_scenario 21)]
if (ticks > 8214 and ticks <= 8396 and ticks mod 2 = 0)[set ts__kal workID + (item
↪  \%3inc table:get poll_scenario 22)]
if (ticks > 8396 and ticks <= 8580 and ticks mod 2 = 0)[set ts__kal workID + (item
↪  \%3inc table:get poll_scenario 23)]
if (ticks > 8580 and ticks <= 8764 and ticks mod 2 = 0)[set ts__kal workID + (item
↪  \%3inc table:get poll_scenario 24)]

if (ticks > 4382 and ticks <= 4562 and (ticks + 1) mod 2 = 0)[set ts__kal homeID +
↪  (item \%3inc table:get poll_scenario 1)]
```

```
if (ticks > 4562 and ticks <= 4744 and (ticks + 1) mod 2 = 0)[set ts__kal homeID +
↪  (item \%3inc table:get poll_scenario 2)]
if (ticks > 4744 and ticks <= 4850 and (ticks + 1) mod 2 = 0)[set ts__kal homeID +
↪  (item \%3inc table:get poll_scenario 3)]
if (ticks > 4850 and ticks <= 5112 and (ticks + 1) mod 2 = 0)[set ts__kal homeID +
↪  (item \%3inc table:get poll_scenario 4)]
if (ticks > 5112 and ticks <= 5292 and (ticks + 1) mod 2 = 0)[set ts__kal homeID +
↪  (item \%3inc table:get poll_scenario 5)]
if (ticks > 5292 and ticks <= 5474 and (ticks + 1) mod 2 = 0)[set ts__kal homeID +
↪  (item \%3inc table:get poll_scenario 6)]
if (ticks > 5474 and ticks <= 5658 and (ticks + 1) mod 2 = 0)[set ts__kal homeID +
↪  (item \%3inc table:get poll_scenario 7)]
if (ticks > 5658 and ticks <= 5842 and (ticks + 1) mod 2 = 0)[set ts__kal homeID +
↪  (item \%3inc table:get poll_scenario 8)]
if (ticks > 5842 and ticks <= 6024 and (ticks + 1) mod 2 = 0)[set ts__kal homeID +
↪  (item \%3inc table:get poll_scenario 9)]
if (ticks > 6024 and ticks <= 6206 and (ticks + 1) mod 2 = 0)[set ts__kal homeID +
↪  (item \%3inc table:get poll_scenario 10)]
if (ticks > 6206 and ticks <= 6390 and (ticks + 1) mod 2 = 0)[set ts__kal homeID +
↪  (item \%3inc table:get poll_scenario 11)]
if (ticks > 6390 and ticks <= 6574 and (ticks + 1) mod 2 = 0)[set ts__kal homeID +
↪  (item \%3inc table:get poll_scenario 12)]
if (ticks > 6574 and ticks <= 6754 and (ticks + 1) mod 2 = 0)[set ts__kal homeID +
↪  (item \%3inc table:get poll_scenario 13)]
if (ticks > 6754 and ticks <= 6936 and (ticks + 1) mod 2 = 0)[set ts__kal homeID +
↪  (item \%3inc table:get poll_scenario 14)]
if (ticks > 6936 and ticks <= 7120 and (ticks + 1) mod 2 = 0)[set ts__kal homeID +
↪  (item \%3inc table:get poll_scenario 15)]
if (ticks > 7120 and ticks <= 7304 and (ticks + 1) mod 2 = 0)[set ts__kal homeID +
↪  (item \%3inc table:get poll_scenario 16)]
if (ticks > 7304 and ticks <= 7484 and (ticks + 1) mod 2 = 0)[set ts__kal homeID +
↪  (item \%3inc table:get poll_scenario 17)]
if (ticks > 7484 and ticks <= 7666 and (ticks + 1) mod 2 = 0)[set ts__kal homeID +
↪  (item \%3inc table:get poll_scenario 18)]
if (ticks > 7666 and ticks <= 7850 and (ticks + 1) mod 2 = 0)[set ts__kal homeID +
↪  (item \%3inc table:get poll_scenario 19)]
if (ticks > 7850 and ticks <= 8034 and (ticks + 1) mod 2 = 0)[set ts__kal homeID +
↪  (item \%3inc table:get poll_scenario 20)]
if (ticks > 8034 and ticks <= 8214 and (ticks + 1) mod 2 = 0)[set ts__kal homeID +
↪  (item \%3inc table:get poll_scenario 21)]
if (ticks > 8214 and ticks <= 8396 and (ticks + 1) mod 2 = 0)[set ts__kal homeID +
↪  (item \%3inc table:get poll_scenario 22)]
```

```
if (ticks > 8396 and ticks <= 8580 and (ticks + 1) mod 2 = 0)[set ts__kal homeID +
↪  (item \%3inc table:get poll_scenario 23)]
if (ticks > 8580 and ticks <= 8764 and (ticks + 1) mod 2 = 0)[set ts__kal homeID +
↪  (item \%3inc table:get poll_scenario 24)]
end

#;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;
to landprice-change
  ask patches with [land10 >= 0][set land10 (random-float .1 + land10)]
end

#;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;
to gu-plot
  set-current-plot "District level"
  set-current-plot-pen "risky"     plot ((count people with [color = violet + 2 and
  ↪  destinationName != "others"]) /
                                        (count people with [destinationName !=
                                        ↪  "others"]) * 100)
  set-current-plot-pen "dangerous" plot ((count people with [color = red and
  ↪  destinationName != "others"]) /
                                        (count people with [ destinationName !=
                                        ↪  "others"]) * 100)
end


#;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;
to dong-plot
  set-current-plot "Subdistrict level"
  set-current-plot-pen "shinsa_risk"    plot((count people with [color = red and
  ↪  districtName = "Shinsa" and destinationName != "others"])   / (count people
  ↪  with [districtName = "Shinsa" and destinationName != "others"])    * 100)
        set-current-plot-pen "nonhyun1_risk"  plot((count people with [color = red
        ↪  and districtName = "Nonhyeon1" and destinationName != "others"])/
        ↪  (count people with [districtName = "Nonhyeon1" and destinationName !=
        ↪  "others"]) * 100)
        set-current-plot-pen "nonhyun2_risk"  plot((count people with [color = red
        ↪  and districtName = "Nonhyeon2" and destinationName != "others"])/
        ↪  (count people with [districtName = "Nonhyeon2" and destinationName !=
        ↪  "others"]) * 100)
```

```
set-current-plot-pen "samsung1_risk"  plot((count people with [color = red
↪  and districtName = "Samseong1" and destinationName != "others"])/
↪  (count people with [districtName = "Samseong1" and destinationName !=
↪  "others"]) * 100)
set-current-plot-pen "samsung2_risk"  plot((count people with [color = red
↪  and districtName = "Samseong2" and destinationName != "others"])/
↪  (count people with [districtName = "Samseong2" and destinationName !=
↪  "others"]) * 100)
set-current-plot-pen "daechi1_risk"   plot((count people with [color = red
↪  and districtName = "Daechi1" and destinationName != "others"]) / (count
↪  people with [districtName = "Daechi1" and destinationName != "others"])
↪  * 100)
set-current-plot-pen "daechi4_risk"   plot((count people with [color = red
↪  and districtName = "Daechi4" and destinationName != "others"]) / (count
↪  people with [districtName = "Daechi4" and destinationName != "others"])
↪  * 100)
set-current-plot-pen "yeoksam1_risk"  plot((count people with [color = red
↪  and districtName = "Yeoksam1" and destinationName != "others"])/ (count
↪  people with [districtName = "Yeoksam1" and destinationName !=
↪  "others"]) * 100)
set-current-plot-pen "yeoksam2_risk"  plot((count people with [color = red
↪  and districtName = "Yeoksam2" and destinationName != "others"])/ (count
↪  people with [districtName = "Yeoksam2" and destinationName !=
↪  "others"]) * 100)
set-current-plot-pen "dogok1_risk"    plot((count people with [color = red
↪  and districtName = "Dogok1" and destinationName != "others"])  / (count
↪  people with [districtName = "Dogok1" and destinationName != "others"])
↪  * 100)
set-current-plot-pen "dogok2_risk"    plot((count people with [color = red
↪  and districtName = "Dogok2" and destinationName != "others"])  / (count
↪  people with [districtName = "Dogok2" and destinationName != "others"])
↪  * 100)
set-current-plot-pen "gaepo1_risk"    plot((count people with [color = red
↪  and districtName = "Gaepo1" and destinationName != "others"])  / (count
↪  people with [districtName = "Gaepo1" and destinationName != "others"])
↪  * 100)
set-current-plot-pen "gaepo4_risk"    plot((count people with [color = red
↪  and districtName = "Gaepo4" and destinationName != "others"])  / (count
↪  people with [districtName = "Gaepo4" and destinationName != "others"])
↪  * 100)
```

```
        set-current-plot-pen "ilwon_risk"     plot((count people with [color = red
        ↪  and districtName = "Ilwonbon" and destinationName != "others"])   /
        ↪  (count people with [districtName = "Ilwonbon" and destinationName !=
        ↪  "others"])    * 100)
        set-current-plot-pen "ilwon1_risk"    plot((count people with [color = red
        ↪  and districtName = "Ilwon1" and destinationName != "others"])  / (count
        ↪  people with [districtName = "Ilwon1" and destinationName != "others"])
        ↪  * 100)
        set-current-plot-pen "ilwon2_risk"    plot((count people with [color = red
        ↪  and districtName = "Ilwon2" and destinationName != "others"])  / (count
        ↪  people with [districtName = "Ilwon2" and destinationName != "others"])
        ↪  * 100)
        set-current-plot-pen "suseo_risk"     plot((count people with [color = red
        ↪  and districtName = "Suseo" and destinationName != "others"])   / (count
        ↪  people with [districtName = "Suseo" and destinationName != "others"])
        ↪  * 100)
        set-current-plot-pen "ap_risk"        plot((count people with [color = red
        ↪  and districtName = "Apgujeong1" and destinationName != "others"])
        ↪  / (count people with [districtName = "Apgujeong1" and destinationName
        ↪  != "others"])       * 100)
        set-current-plot-pen "chungdam_risk"  plot((count people with [color = red
        ↪  and districtName = "Cheongdam" and destinationName != "others"])/
        ↪  (count people with [districtName = "Cheongdam" and destinationName !=
        ↪  "others"]) * 100)
        set-current-plot-pen "daechi2_risk"   plot((count people with [color = red
        ↪  and districtName = "Daechi2" and destinationName != "others"]) / (count
        ↪  people with [districtName = "Daechi2" and destinationName != "others"])
        ↪  * 100)
        set-current-plot-pen "gaepo2_risk"    plot((count people with [color = red
        ↪  and districtName = "Gaepo2" and destinationName != "others"])  / (count
        ↪  people with [districtName = "Gaepo2" and destinationName != "others"])
        ↪  * 100)
        set-current-plot-pen "segok_risk"     plot((count people with [color = red
        ↪  and districtName = "Segok" and destinationName != "others"])   / (count
        ↪  people with [districtName = "Segok" and destinationName != "others"])
        ↪  * 100)
end

to age-plot
  set-current-plot "By Age Group"
```

```
  set-current-plot-pen "Young"  plot(count people with [age < 15 and color = red
  ↪  and destinationName != "others"]) / (count people with [age < 15 and
  ↪  destinationName != "others"]) * 100
  set-current-plot-pen "Middle" plot((count people with [age >= 15 and age < 65 and
  ↪  color = red and destinationName != "others"]) / (count people with [age >= 15
  ↪  and age < 65 and destinationName != "others"]) * 100)
  set-current-plot-pen "Old"    plot((count people with [age >= 65 and color = red
  ↪  and destinationName != "others"]) / (count people with [age >= 65 and
  ↪  destinationName != "others"]) * 100)
end

to edu-plot
  set-current-plot "By Education"
  set-current-plot-pen "High" plot((count people with [edu >= 3 and color = red and
  ↪  destinationName != "others"]) / (count people with [edu >= 3 and
  ↪  destinationName != "others"]) * 100)
  set-current-plot-pen "Low"  plot((count people with [edu < 3 and color = red and
  ↪  destinationName != "others"]) / (count people with [edu < 3 and
  ↪  destinationName != "others"]) * 100)

end

to pm10-plot
  set-current-plot "PM10 patches"
  set-current-plot-pen "pm10-shinsa-road"  plot [ts__kal] of patch 24 253
end
```

B.2 Chapter 5: Simulating vehicles and pedestrians in Seoul CBD

```
extensions [gis csv table nw ]

globals [
  pollution-data
  area
  roads
  streets
  mean-pm10
  pm10-back pm10-road pm10-others
]
```

```
breed [nodes node]
breed [area-labels area-label]
breed [buildings building]
breed [s_entrances s_entrance]
breed [cars car]
breed [employees employee]
breed [drivers driver]

links-own [
  road-name
  is-road?
  max-spd
  Daero?
  weight
]

patches-own [
  is-research-area?
  is-endpoint?
  intersection
  mybuilding
  building_info
  dilution
  countdown
  dong-name
  dong-code
  road_buffer
  pm10
  pm10_indoor
  centroid? #;;is it the centroid of a building?
  id   #;;if it is a centroid of a building, it has an ID that represents the
  ↪  building
  b_entrance #;;nearest vertex on road. only for centroids.
]

nodes-own [
  name
  dong_code
  endpoint?
  line-start
  line-end
```

```
  auto?
  green-light?
  intersection?
  b_entrance?
  dist-original  #;;distance from original point to here
]

s_entrances-own [
 Line #;; Seoul Metro has lines from 1 to 9
]

buildings-own [
 close-to-road?
 #;count-employees
]

cars-own [
  speed
  fueltype
  tyre-wear
  brake-wear
  surface-wear
  total-emission
  origin #;;a vertex. where the vehicle begins the trip
  destination #;; allocated destination
  myoffice
  goal  #;;the b_entrance of the destination on the road
  path-work #;; an agentset containing nodes to visit in the shortest path
  path-home
  nodes-remaining
  current
  to-node
  myroad
  current-link
  district_name district_code
  link-counter   #; counter for distance along path
  direction      #; +1 work to work and -1 for to home
  time-at-work   #; how long to spend at work
  random-car     #; work like venice model, or go home to work and back
  parked         #; parked cars don't interfere with other traffic
  owner
  health
```

```
  leave-home-hour
  leave-home-mins
  unwell_history
  work-near-roads?
]


employees-own [
  origin #;;a vertex. where the vehicle begins the trip
  origin_patch
  myhome
  myoffice
  goal  #;;the b_entrance of the destination on the road
  current
  Heuristic
  arrived?
  time-at-work   #; how long to spend at work
  direction
  arrive-tick
  leave-home-hour
  leave-home-mins
  work-near-roads?
  health
  unwell_history
]

drivers-own [
  commute_method
  my_car
  age
  health
  #;work-near-roads?
]

#;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;
#;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;

to setup
  ca
  set-gis
  add-labels
  activate-links
```

```
  set-signals
  set-random-cars
  set-resident-cars
  set-resident-driver
  set-incoming-traffic
  set-subway-commuters
  set-OD
  set-path-node
  set-path-link
  set-road-ends
  set-pm10
  set-pm10-others
  ask cars with [not random-car] [to-work-setup]
  reset-ticks

end

#;;-------------------------------------------------------------
to go
  set Jongno_p     precision ([pm10] of patch 134 98) 2
  set Jung_p       precision ([pm10] of patch 38 68) 2
  set JongnoKerb_p precision ([pm10] of patch 105 95) 2
  set Samil_p      precision ([pm10] of patch 77 77) 2 #;"Samil-daero"
  set Sejong_p     precision ([pm10] of patch 43 112) 2 #;"Sejong-daero"
  set Pirum_p      precision ([pm10] of patch 22 148) 2 #;"Pirum-daero"
  set Yulgok_p     precision ([pm10] of patch 124 117) 2 #;"Daehak-ro"
  set Drivers_p    precision ((count cars with [not random-car and health < 100] /
                               count cars with [not random-car]) * 100) 3
  set Walkers_p    precision ((count(employees with [health < 100]) / count
  ↪  employees) * 100) 3
  set mean-pm10    precision mean [pm10] of patches with [is-research-area? = true]
  ↪  3

  if (ticks + 1) >= 127740 [stop]
  ask cars [
    let is-weekend? item 6 table:get pm10-road (ticks + 1)
    let what-time?  item 1 table:get pm10-road (ticks + 1)
     let hours item 1 table:get pm10-back (ticks + 1)
    let minutes item 4 table:get pm10-back (ticks + 1)
    let weekday? item 5 table:get pm10-back (ticks + 1)
    let travel-hours what-time? >= (8 + random 2) and what-time? < 22
    ifelse (random-car)[move speed]
```

```
      [ if is-weekend? = false and weekday? != "Mon" [travel speed]
        if is-weekend? = false and weekday? = "Mon" and hours = 6 and minutes = 59
        ↪  [to-work-setup set time-at-work 540 + random 61]
        if is-weekend? = false and weekday? = "Mon" and hours >= 7  [travel speed]
        if is-weekend? = false and weekday? = "Mon" and hours >= 7 and minutes = 5
        ↪  and parked [park]
        if (is-weekend? = true and awareness = "no" and travel-hours) [move speed]
        ↪  #;; move resident cars on weekends only when awareness off
        if (is-weekend? = true and awareness = "yes" and travel-hours and health >=
        ↪  100) [move speed] #;; take rest when awareness on
        if (is-weekend? = true and awareness = "yes" and travel-hours and health <
        ↪  100)  [move-to origin to-work-setup] #;; take rest when awareness on
        if (is-weekend? = true and what-time? >= 23) [move-to origin to-work-setup]
        ↪  ] #;; Flying cars may appear.
                                                                  #;; They are just
                                                                  ↪  heading home
                                                                  ↪  without using
                                                                  ↪  the road
                                                                  ↪  links.
  ]
  speed-up
  set-signal-colours
  meet-traffic-lights
  drive-out-of-cbd
  add-cars
  pollute
  fadeout
  kill-cars
  add-employees
  move-employees
  move-drivers
  health-loss
  health-recovery
  validation-plot
  age-plot
  tick


end


#;;;
```

```
#;;;
```

```
        set centroid? true
        set building_info self
        set id gis:property-value feature "ID"
        ifelse [road_buffer] of patch-here = true [set close-to-road? true][set
        ↪  close-to-road? false]
    ]]]
  ask patches with [centroid? != true][set centroid? false set ID "not given"  ]


#;; Create subway s_entrances
  let subway gis:load-dataset "GIS/Subway_SeoulCBD.shp"
  foreach gis:feature-list-of subway [ vector-feature ->
    foreach gis:vertex-lists-of vector-feature [ vertex ->
      foreach vertex [ point ->
        let location gis:location-of point
        if not empty? location [
         create-s_entrances 1  [
            set xcor item 0 location
            set ycor item 1 location
            set shape "flag"
            set size 3
            set Line gis:property-value vector-feature "LineNo"
            if Line = 1 [ set color blue ]
            if Line = 2 [ set color lime ]
            if Line = 3 [ set color orange ]
            if Line = 4 [ set color sky ]
            if Line = 5 [ set color magenta ]
  ]]]]]

  ask s_entrances [
    if count s_entrances-here > 1 [
      ask other s_entrances-here [ die ] ]]

#;; Create turtles representing the nodes. Create links to connect them
  foreach gis:feature-list-of roads [ vector-feature ->
    let first-vertex gis:property-value vector-feature "UP_FROM_NO"
    let last-vertex gis:property-value vector-feature "UP_TO_NO"

    foreach  gis:vertex-lists-of vector-feature [ vertex ->
      let previous-node nobody
      foreach vertex [ point ->
        let location gis:location-of point
```

```
        if not empty? location
        [ create-nodes 1 [
            set xcor item 0 location
            set ycor item 1 location
            set size 0.05
            set shape "circle"
            set color one-of base-colors
            set hidden? false
            set line-start first-vertex
            set line-end last-vertex
            ifelse previous-node = nobody
              []
              [create-link-with previous-node] #; create link to previous node
               set previous-node self]
        ]
  ] #; end of foreach vertex
  ] #; end of foreach  gis:vertex-lists-of vector-feature
  ] #; end of foreach gis:feature-list-of roads


  #;;delete duplicate vertices
  #;;(there may be more than one vertice on the same patch due to reducing size of
  ↪  the map).
  #;;therefore, this map is simplified from the original map.
    ask nodes [
      if count nodes-here > 1 [
      ask other nodes-here  [
        ask myself [create-links-with other [link-neighbors] of myself]
        die]]
     ]

  #;;find nearest node to become b_entrance
  ask patches with [centroid? = true][
    set b_entrance min-one-of nodes in-radius 200 [distance myself]
    ask b_entrance [set b_entrance? true]]
  ask patches with [centroid? = false][ask nodes [set b_entrance? false]]

end

to draw-4daemoon
  gis:set-drawing-color [ 229 255 204]    gis:fill area 0 #;;RGB color
  gis:set-drawing-color [  64  64  64]    gis:draw area 1
```

```
end

to draw-map
  import-drawing "GIS/map.png"
end

to add-labels
foreach gis:feature-list-of area [vector-feature ->
    let centroid gis:location-of gis:centroid-of vector-feature
       if not empty? centroid
      [ create-area-labels 1
        [ set xcor item 0 centroid
          set ycor item 1 centroid
          set size 0
          set label-color blue
          set label gis:property-value vector-feature "adm_dr_nm_"
      ]]]
  ask nodes [set dong_code [dong-code] of patch-here]

end

to activate-links
  ask links [
    set is-road? true
    let way list [line-start] of end1 [line-end] of end2
    let daero ["Jahamun-ro"  "Sajik-ro"  "Samil-daero"  "Yulgok-ro"  "Toegye-ro"
    ↪  "Saemunan-ro 3-gil" "Jangchungdan-ro"
      "Taepyeong-ro"  "Sejong-daero"  "Jong-ro"  "Eulji-ro"  "Seosomun-ro"
      ↪  "Donhwamun-ro" "Sejong-daero 23-gil"]

    foreach gis:feature-list-of roads [ vector-feature-sub ->
      let mspeed gis:property-value vector-feature-sub "MAX_SPD"
      let vector-start gis:property-value vector-feature-sub "UP_FROM_NO"
      let vector-end gis:property-value vector-feature-sub "UP_TO_NO"
      let start-end list vector-start vector-end
      let end-start list vector-end vector-start

      if way = start-end [set road-name gis:property-value vector-feature-sub
      ↪  "ROAD_NAME_"]
      if road-name = one-of daero [set Daero? true]
      if road-name = 0 or road-name = "" [set road-name [name] of end2 ]
      set max-spd read-from-string mspeed
```

```
        ]
  ]
end

to set-signals
#; Set Signals in junctions
  ask nodes [
    ifelse count my-links > 2
      [set size 3
       set intersection? true
       set auto? random 11
    ]
      [#;set hidden? true
       set color grey
      ]
  ]

  ask nodes with [intersection? = true][
    if auto? >= 5 [set color green set green-light? true]
    if auto? <  5 [set color red set green-light? false]
    set pcolor black + 5
    ask neighbors [set pcolor black + 4]
    ask patches with [pcolor = black + 4] [set intersection true]
  ]
end

to set-road-ends
  let endpoints0 map node [5465 5463 5464 5451 5452 5798 5797 5685 5690 6745 6753
  ↪  5653 5651 5652 6687 6679] #;6778 6704 ]
  let endpoints filter is-node? endpoints0

 foreach endpoints [ep ->
    ask ep [set endpoint? true ask patch-here [set is-endpoint? true ]]
]
  ask nodes with [endpoint? != true][set endpoint? false]
  ask patches with [is-endpoint? = true][ask neighbors [set is-endpoint? true]]
  ask patches with [is-endpoint? != true][set is-endpoint? false]
end



to set-random-cars
  set-default-shape cars "car"
```

```
  create-cars 50 [ #;; random-cars
    set parked false
    set random-car true
    set size 2
    set speed 0
    set destination nobody
    set origin one-of nodes move-to origin
            ]
  ask n-of (int(.7 * count cars)) cars [set fueltype "Gasoline"]
  ask cars with [fueltype != "Gasoline"][set fueltype "Diesel" ]
end

to set-path-node
    ask cars [
    #; Randomly choose a target node to walk to
    let target [goal] of self
    #;print target
    if target != nobody [
      #; Remember the starting node
      set current one-of nodes-here
      #; Define a path variable from the current node- take all but
      #; the first item (as first item is current node)
      let path nobody
      ask links [ set weight link-length ]
      ask current [
        set path but-first nw:turtles-on-weighted-path-to target weight
      ]

      #; Indicate the end node
      ask last path [
        set color [color] of self
        set b_entrance? true
        set size 0.5 ]

      let path-work0 lput path path-work #;; assign all the nodes that leads to the
      ↪  destination node
      set path-work item 0 path-work0
      set path-work fput origin path-work
      set to-node first path #;; or can code as ---> item 0 item 0 path-work
    ]

    set nodes-remaining length [path-work] of self
```

```
    set current-link link [who] of [current] of self [who] of [to-node] of self
    face to-node
    let points n-values [nodes-remaining] of self [x -> ([nodes-remaining] of self
    ↪ - 1) - x]
    foreach points [ x ->
    set path-home lput item x [path-work] of self path-home
    ]
    set path-home lput origin path-home
   ]
end

to set-path-link
  ask cars [
    let vertex0 path-work
    let vertex1 remove-item 0 path-work
    let imsi0 []
    foreach vertex0 [ x ->
      let element [who] of x
      set imsi0 lput element imsi0
    ]
    let imsi00 bl imsi0
    let imsi1 []
    foreach vertex1 [ x ->
      let element [who] of x
      set imsi1 lput element imsi1
    ]
   set myroad (map link imsi00 imsi1)
  ]
end

to set-pm10
  #; Import daily pollution
  let p0 csv:from-file "GIS/jongno_pm10.csv"
  let poll-value remove-item 0 p0  #;;remove headers in the csv file
  let rep 0  #;; loop
  set pm10-back table:make
  set pm10-road table:make

  foreach poll-value [poll ->
    if item 1 poll = "Back" [
    let counter item 0 poll #;; counter
    let date/hour list (item 2 poll)(item 3 poll) #;; add date and place
```

```
    let value lput item 4 poll date/hour
    let pm10_ lput item 5 poll value
    let minute lput item 6 poll pm10_
    let wdays lput item 7 poll minute
    let isweekend lput item 8 poll wdays
    table:put pm10-back counter isweekend
  ]
    if item 1 poll = "Road" [
    let counter item 0 poll #;; counter
    let date/hour list (item 2 poll)(item 3 poll) #;; add date and place
    let value lput item 4 poll date/hour
    let pm10_ lput item 5 poll value
    let minute lput item 6 poll pm10_
    let wdays lput item 7 poll minute
    let isweekend lput item 8 poll wdays
    table:put pm10-road counter isweekend
  ]
  ]
  set rep rep + 1

  ask patches with [is-research-area? = true and not road_buffer]
    [set pm10 (item 2 table:get pm10-back 1) + random-float (item 3 table:get
    ↪  pm10-back 1) set pm10_indoor nobody]

  ask patches with [is-research-area? = true and road_buffer = true]
    [set pm10 (item 2 table:get pm10-road 1) + random-float (item 3 table:get
    ↪  pm10-road 1)]
end


to set-pm10-others
 #; Import daily pollution
  let p0 csv:from-file "GIS/other_pm10.csv"
  let poll-value remove-item 0 p0  #;;remove headers in the csv file
  let rep 0  #;; loop

  set pm10-others table:make

  foreach poll-value [poll ->
    if item 1 poll = "Mix" [
    let counter item 0 poll #;; counter
    let date/hour list (item 2 poll)(item 3 poll) #;; add date and place
```

```
    let value lput item 4 poll date/hour
    let #sd lput item 5 poll value
    let minute lput item 6 poll #sd
    let wdays lput item 7 poll minute
    let isweekend lput item 8 poll wdays
    table:put pm10-others counter isweekend
  ]]

  ask patch max-pxcor max-pycor
  [set pm10 (item 2 table:get pm10-others 1) + random-float (item 3 table:get
  ↪  pm10-others 1)]
end



to set-resident-cars
  let vehicle csv:from-file "GIS/Seoul_Vehicle_sample.csv"
  let rawheader item 0 vehicle
  let carTT remove-item 0 rawheader
  let carstat remove-item 0 vehicle

  let districtCar table:make
  let districtadminCode table:make

  foreach carstat [ code ->
    let gasdiesel list (item 2 code)(item 3 code)
    let total lput item 4 code gasdiesel
    table:put districtCar item 0 code total
    table:put districtadminCode item 0 code item 1 code
  ]
    foreach table:keys districtCar [ v ->
    let carGroupID 0
    foreach table:get districtCar v [ xx ->
      create-cars xx [ #;; resident car ratio
        set random-car false
        set time-at-work 540 + random 61
        setupCarGroup carGroupID
        set size 2
        set district_code v
        set district_name table:get districtadminCode v
        set shape "car"
        set origin one-of nodes with [dong_code = [district_code] of myself]
```

```
        set destination nobody
        set health 300
        set unwell_history false
        set leave-home-hour 7
        set leave-home-mins random 40
      ]
      set carGroupID carGroupID + 1
    ]]

  ask cars [ifelse origin != nobody [move-to origin]
    [set origin one-of nodes with [dong_code = 1101054]
    move-to origin
    ]
  ]
end

to setupCarGroup [Car_ID]
  if Car_ID = 0 [set fueltype "Gasoline" set color green]
  if Car_ID = 1 [set fueltype "Diesel" set color black + 1]
  if Car_ID = 2 [set fueltype "LPG" set color blue + 2]
end

to set-OD   #;; Decomposing matrix
  let odcar csv:from-file "GIS/od_car.csv"
  let rawheader item 0 odcar
  let destinationNames remove-item 0 rawheader
  let ODMat remove-item 0 odcar

  let ODMatrix table:make
  foreach ODMat [ origin-chart ->
    let number remove-item 0 origin-chart
    table:put ODMatrix item 0 origin-chart number
  ]

  foreach table:keys ODMatrix [originName ->
    let matrix-loop 0
    let Num count cars with [district_code = originName]
    let totalUsed 0
    let number 0

  foreach table:get ODMatrix originName [ percent ->
      let newDestination item matrix-loop destinationNames
```

```
      ifelse (newDestination != 1102060)
          [set number precision (percent * Num) 0 set totalUsed totalUsed + number]
          [set number Num - totalUsed]

      let carsRemaining cars with [district_code = originName and destination =
      ↪  nobody]

      ask n-of number carsRemaining [
        set destination one-of patches with [dong-code = newDestination and
        ↪  centroid? = true and intersection != true]
        set myoffice [building_info] of destination
        set goal [b_entrance] of destination
        ifelse [close-to-road?] of myoffice = true [ set work-near-roads? true ][
        ↪  set work-near-roads? false ]
        set path-work []
        set path-home []
        set myroad []

        while [goal = origin] [
          set destination one-of patches with [dong-code = newDestination and
          ↪  centroid? = true and intersection != true]
          set goal [b_entrance] of destination
          set myoffice [building_info] of destination
          ifelse [close-to-road?] of myoffice = true [ set work-near-roads? true ][
          ↪  set work-near-roads? false ]
        ]
       ]

     set matrix-loop matrix-loop + 1
    ]
output-type totalused output-type " " output-type Num output-type " " output-print
↪  originName
  ]

  #;;;;; Kill cars for the time being
  ask cars with [destination = nobody][die]
  type (word "Cars not able to find their destinations were killed for the time
  ↪  being" "\n"
    "Flying cars might appear on Sat&Sun Nights: Simplifying the 'head home'
    ↪  procedure after a day trip")
  #;;;; Change vehicle's goal if loaded at endpoints
```

```
  ask cars with [origin = [endpoint?] of nodes ][move-to one-of nodes in-radius 10
  ↪  with [endpoint? = false] set endpoint? nodes-here]
  ask cars [
    if goal = [endpoint?] of nodes [ set goal one-of nodes with [endpoint? = false]
  ]]
end

to to-work-setup
  #;initialisation
  set parked false
  set to-node goal
  set current-link item 0 myroad
  ifelse  ([end1] of current-link = origin) [set to-node [end2] of current-link]
                                            [set to-node [end1] of current-link]
  face to-node
  set link-counter 0
  set direction 1
end


to to-home-setup
  #;initialisation
  set parked false
  set to-node origin
  set current-link last myroad
  ifelse  ([end1] of current-link = goal)   [set to-node [end2] of current-link]
                                            [set to-node [end1] of current-link]
  face to-node
  set link-counter (length myroad) - 1
  set direction -1
end


to set-resident-driver
  ask cars with [not random-car] [
    let c self
    hatch-drivers 1 [
     set shape "person business"
     set commute_method "Drive"
     set my_car c
     set age 25 + random 60
     set health 300
```

```
      #;set work-near-roads? [work-near-roads?] of c
  ]

  let my_driver item 0 [self] of drivers-on patch-here
  if [my_car] of my_driver = c [set owner my_driver]]
  #;ask drivers with [work-near-roads? != true][set work-near-roads? false]
  ask drivers [set hidden? true ]
end

to set-subway-commuters
  create-employees no-of-employees [
    set size 2
    set shape "person business"
    set arrived? false
    set origin one-of s_entrances
    set origin_patch [patch-here] of origin
    set myhome (patch max-pxcor max-pycor)
    set goal nobody
    set Heuristic 0
    set current 0
    set direction 0
    set arrive-tick 0
    set leave-home-hour nobody
    set leave-home-mins random 60
    set health 300
    ]

  ask n-of (count employees * .9) employees [
    set leave-home-hour (6 + random 3)
  ]

  ask employees
  [
    move-to patch max-pxcor max-pycor
    if leave-home-hour = nobody [ set leave-home-hour (8 + random 4) ]
    let p 0
    ask origin [set p one-of buildings in-radius 20 with [ is-research-area? =
    ↪  true]]
    set myoffice p
    set goal [patch-here] of p
    set work-near-roads? [close-to-road?] of myoffice = true
  ]
```

```
end


to set-incoming-traffic
  let p0 csv:from-file "GIS/traffic_calibration.csv"
  let traff-value remove-item 0 p0  #;;remove headers in the csv file

  let t_sajik table:make
  let t_jahamoon table:make
  let t_daesagwan table:make
  let t_daehak table:make
  let t_jongno table:make
  let t_dongho table:make
  let t_ns2ho table:make
  let t_soparo table:make
  let t_saemunan table:make
  let t_ssm table:make

  foreach traff-value [traff ->
    if item 1 traff = "A-02" [
    let counter item 0 traff #;; counter
    let date/hour list (item 2 traff)(item 3 traff) #;; add date and place
    let #count lput item 4 traff date/hour
    let weekday lput item 5 traff #count
    table:put t_sajik counter weekday
  ]

  if item 1 traff = "A-03" [
    let counter item 0 traff #;; counter
    let date/hour list (item 2 traff)(item 3 traff) #;; add date and place
    let #count lput item 4 traff date/hour
    let weekday lput item 5 traff #count
    table:put t_jahamoon counter weekday
  ]

  if item 1 traff = "A-04" [
    let counter item 0 traff #;; counter
    let date/hour list (item 2 traff)(item 3 traff) #;; add date and place
    let #count lput item 4 traff date/hour
    let weekday lput item 5 traff #count
    table:put t_daesagwan counter weekday
```

```
]

if item 1 traff = "A-07" [
  let counter item 0 traff #;; counter
  let date/hour list (item 2 traff)(item 3 traff) #;; add date and place
  let #count lput item 4 traff date/hour
  let weekday lput item 5 traff #count
  table:put t_daehak counter weekday
]

if item 1 traff = "A-08" [
  let counter item 0 traff #;; counter
  let date/hour list (item 2 traff)(item 3 traff) #;; add date and place
  let #count lput item 4 traff date/hour
  let weekday lput item 5 traff #count
  table:put t_jongno counter weekday
]

if item 1 traff = "A-10" [
  let counter item 0 traff #;; counter
  let date/hour list (item 2 traff)(item 3 traff) #;; add date and place
  let #count lput item 4 traff date/hour
  let weekday lput item 5 traff #count
  table:put t_dongho counter weekday
]

if item 1 traff = "A-21" [
  let counter item 0 traff #;; counter
  let date/hour list (item 2 traff)(item 3 traff) #;; add date and place
  let #count lput item 4 traff date/hour
  let weekday lput item 5 traff #count
  table:put t_ns2ho counter weekday
]

if item 1 traff = "A-24" [
  let counter item 0 traff #;; counter
  let date/hour list (item 2 traff)(item 3 traff) #;; add date and place
  let #count lput item 4 traff date/hour
  let weekday lput item 5 traff #count
  table:put t_soparo counter weekday
]
```

```
  if item 1 traff = "A-14" [
    let counter item 0 traff #;; counter
    let date/hour list (item 2 traff)(item 3 traff) #;; add date and place
    let #count lput item 4 traff date/hour
    let weekday lput item 5 traff #count
    table:put t_saemunan counter weekday
  ]

  if item 1 traff = "A-16" [
    let counter item 0 traff #;; counter
    let date/hour list (item 2 traff)(item 3 traff) #;; add date and place
    let #count lput item 4 traff date/hour
    let weekday lput item 5 traff #count
    table:put t_ssm counter weekday
  ]
  ]

extra-setting
end


to extra-setting
 set v_sajik []
 let k_sajik table:keys t_sajik
 foreach k_sajik
      [triggers -> repeat 60 [
          let carcounts item 2 table:get t_sajik triggers
          set v_sajik lput carcounts  v_sajik ]]

 set v_jahamoon []
 let k_jahamoon table:keys t_jahamoon
 foreach k_jahamoon
      [triggers -> repeat 60 [
          let carcounts item 2 table:get t_jahamoon triggers
          set v_jahamoon lput carcounts  v_jahamoon ]]

 set v_daesagwan []
 let k_daesagwan table:keys t_daesagwan
 foreach k_daesagwan
      [triggers -> repeat 60 [
          let carcounts item 2 table:get t_daesagwan triggers
          set v_daesagwan lput carcounts v_daesagwan ]]
```

```
set v_daehak []
let k_daehak table:keys t_daehak
foreach k_daehak
     [triggers -> repeat 60 [
         let carcounts item 2 table:get t_daehak triggers
         set v_daehak lput carcounts v_daehak ]]

set v_jongno []
let k_jongno table:keys t_jongno
foreach k_jongno
     [triggers -> repeat 60 [
         let carcounts item 2 table:get t_jongno triggers
         set v_jongno lput carcounts v_jongno ]]

set v_dongho []
let k_dongho table:keys t_dongho
foreach k_dongho
     [triggers -> repeat 60 [
         let carcounts item 2 table:get t_dongho triggers
         set v_dongho lput carcounts v_dongho ]]

set v_ns2ho []
let k_ns2ho table:keys t_ns2ho
foreach k_ns2ho
     [triggers -> repeat 60 [
         let carcounts item 2 table:get t_ns2ho triggers
         set v_ns2ho lput carcounts v_ns2ho ]]

set v_soparo []
let k_soparo table:keys t_soparo
foreach k_soparo
     [triggers -> repeat 60 [
         let carcounts item 2 table:get t_soparo triggers
         set v_soparo lput carcounts v_soparo ]]

set v_saemunan []
let k_saemunan table:keys t_saemunan
foreach k_saemunan
     [triggers -> repeat 60 [
         let carcounts item 2 table:get t_saemunan triggers
         set v_saemunan lput carcounts v_saemunan ]]
```

```
 set v_ssm []
 let k_ssm table:keys t_ssm
 foreach k_ssm
      [triggers -> repeat 60 [
          let carcounts item 2 table:get t_ssm triggers
          set v_ssm lput carcounts v_ssm ]]
end


#;;;;;;;;;;;;;;;;;;;;;;
#;;;;;;;;;;GO#;;;;;;;;;;
#;;;;;;;;;;;;;;;;;;;;;;
to move [dist] #;;
  set current one-of nodes-here
  let dxnode distance to-node
  ifelse dxnode > dist [forward dist] [
    let nextlinks [my-links] of to-node

    ifelse count nextlinks = 1
    [ set-next-car-link current-link to-node ]
    [ set-next-car-link  one-of nextlinks with [self != [current-link] of myself]
    ↪  to-node]
    move dist - dxnode
    set-emission
   ]
end

to set-next-car-link [way n]
  set current-link way
  move-to n
  ifelse n = [end1] of way [set to-node [end2] of way] [set to-node [end1] of way]
  face to-node
end

to travel [dist]
  set current one-of nodes-here    #;; if "current" isn't assigned, then the
  ↪  vehicles will fly everywhere..
  let dxnode distance to-node

  ifelse dxnode > dist
    [forward dist]
```

```
    [
    move-to to-node
    ifelse (direction = 1  and to-node != goal) or (direction = -1  and to-node !=
    ↪  origin)
     [
       set link-counter link-counter + direction
       set current-link item link-counter myroad
       ifelse  ([end1] of current-link = to-node)   [set to-node [end2] of
       ↪  current-link]
                                                    [set to-node [end1] of
                                                    ↪  current-link]
       face to-node
       travel dist - dxnode
       set-emission
     ]
     [ set speed 0
       park
     ]
    ]
end

to park
  let hours item 1 table:get pm10-back (ticks + 1)
  let minutes item 4 table:get pm10-back (ticks + 1)
  let is-weekend? item 6 table:get pm10-back (ticks + 1)

  set parked true
  set tyre-wear  0
  set brake-wear 0
  set surface-wear 0
  set total-emission 0

  ifelse to-node = goal or current = goal
    [ set time-at-work time-at-work - 1
      if (time-at-work <= 0) [to-home-setup ]
     ]
    [ if (hours = leave-home-hour and minutes = leave-home-mins) and is-weekend? =
    ↪  false
         [to-work-setup set time-at-work 540 + random 61]  #;; cars will go to work
         ↪  after 7am
    ]
end
```

```
to set-emission
  #;; Tyre wear
  let velocity-tyre 0
  let EF1# 0
  let random_no# random 2
  ifelse speed < 2 [set velocity-tyre 1.39][set velocity-tyre (-.00974 * speed) +
  ↪ 1.78]
  ifelse random_no# = 0 [set EF1# (0.0107 - random-float .004)][set EF1# (0.0107 +
  ↪ random-float .006)]
  set tyre-wear (emission-factor * EF1# * .6 * velocity-tyre) * 333

  #;; Brake wear
  let velocity-brake 0
  let EF2# 0
  let random_no1# random 2
  ifelse speed < 2 [set velocity-brake 1.67][set velocity-brake (-.0270 * speed) +
  ↪ 2.75]
  ifelse random_no1# = 0 [set EF2# (0.0067 - random-float .004)][set EF2# (0.0067 +
  ↪ random-float .006)]
  set brake-wear (emission-factor * EF2# * .6 * velocity-brake) * 333

  #;;  Surface wear
  set surface-wear (emission-factor * .015 * .5) * 333

  #;; Total Emission
  let dilution# (0.5 + random-float .50) #;; wind speed, turbulance, traffic
  ↪ intensity
  set total-emission (tyre-wear + brake-wear + surface-wear) * dilution#
end

to speed-up
  let max-speed 3 + random-float 2
  let min-speed .5
  ask cars [
    set speed min-speed + random-float (max-speed - min-speed)
    let car-ahead one-of (cars-on patch-ahead 1) with [heading = [heading] of
    ↪ myself]
    ifelse car-ahead != nobody and not [parked] of car-ahead
      [ slow-down car-ahead] [set speed speed]
  ]
```

```
end

to slow-down [car-ahead ]
  #;; slow down so you are driving more slowly than the car ahead of you
  let deceleration 0 + random-float 3
  set speed [ speed ] of car-ahead - deceleration

end

to set-signal-colours
  let hours item 1 table:get pm10-back (ticks + 1)

  ask nodes with [intersection? = true] [
   ifelse hours >= 2 and hours < 6
    [ set auto? 10 set color green set green-light? true ]
    [ set auto? auto? - 1
      if auto? >= 5 [set color green set green-light? true]
      if auto? <  5 [set color red set green-light? false]
      if auto? <= 0 [set auto? 5 + random 6]]
  ]
end

to meet-traffic-lights
  ask cars [
  if any? nodes in-radius 2 with [green-light? = false][ set speed 0 ]
  if any? other nodes in-radius 1 with [green-light? = true]
    [set speed (speed + (.5 + random-float 1)) ]
  ]
end

to add-cars
  let CBD_Entrance []
  let input_location (map patch [14 58 131 149 140 102 63 38 18 2 154] [181 184 164
  ↪  99 46 17 34 40 86 103 191])
  foreach input_location [ x ->
    ask x [
      let val [who] of min-n-of 2 nodes [distance myself]
      let val-link link (item 0 val) (item 1 val)
      if val-link != nobody [set CBD_Entrance lput val-link CBD_Entrance]
      ]]

  create-cars round (item ticks v_sajik * car_ratio / 60) [
```

```
  car-info
  let l item 0 CBD_Entrance
  ifelse l = item 0 CBD_Entrance or l = item 1 CBD_Entrance or
         l = item 4 CBD_Entrance or l = item 5 CBD_Entrance or l = item 6
         ↪ CBD_Entrance
     [set-next-car-link l [end2] of l] #;; ask the vehicles to head south
     [set-next-car-link l [end1] of l] #;; ask the vehicles to head north
]

create-cars round (item ticks v_jahamoon * car_ratio / 60) [
  car-info
  let l item 1 CBD_Entrance
  ifelse l = item 0 CBD_Entrance or l = item 1 CBD_Entrance or
         l = item 4 CBD_Entrance or l = item 5 CBD_Entrance or l = item 6
         ↪ CBD_Entrance
     [set-next-car-link l [end2] of l] #;; ask the vehicles to head south
     [set-next-car-link l [end1] of l] #;; ask the vehicles to head north
]

create-cars round (item ticks v_daesagwan * car_ratio / 60) [
  car-info
  let l item 2 CBD_Entrance
  ifelse l = item 0 CBD_Entrance or l = item 1 CBD_Entrance or
         l = item 4 CBD_Entrance or l = item 5 CBD_Entrance or l = item 6
         ↪ CBD_Entrance
     [set-next-car-link l [end2] of l] #;; ask the vehicles to head south
     [set-next-car-link l [end1] of l] #;; ask the vehicles to head north
]

create-cars round (item ticks v_daehak * car_ratio / 60) [
  car-info
  let l item 3 CBD_Entrance
  ifelse l = item 0 CBD_Entrance or l = item 1 CBD_Entrance or
         l = item 4 CBD_Entrance or l = item 5 CBD_Entrance or l = item 6
         ↪ CBD_Entrance
     [set-next-car-link l [end2] of l] #;; ask the vehicles to head south
     [set-next-car-link l [end1] of l] #;; ask the vehicles to head north
]

create-cars round (item ticks v_jongno * car_ratio / 60) [
  car-info
  let l item 4 CBD_Entrance
```

```
 ifelse l = item 0 CBD_Entrance or l = item 1 CBD_Entrance or
        l = item 4 CBD_Entrance or l = item 5 CBD_Entrance or l = item 6
        ↪ CBD_Entrance
    [set-next-car-link l [end2] of l] #;; ask the vehicles to head south
    [set-next-car-link l [end1] of l] #;; ask the vehicles to head north
]

create-cars round (item ticks v_dongho * car_ratio / 60) [
 car-info
 let l item 5 CBD_Entrance
 ifelse l = item 0 CBD_Entrance or l = item 1 CBD_Entrance or
        l = item 4 CBD_Entrance or l = item 5 CBD_Entrance or l = item 6
        ↪ CBD_Entrance
    [set-next-car-link l [end2] of l] #;; ask the vehicles to head south
    [set-next-car-link l [end1] of l] #;; ask the vehicles to head north
]

create-cars round (item ticks v_ns2ho * car_ratio / 60) [
 car-info
 let l item 6 CBD_Entrance
 ifelse l = item 0 CBD_Entrance or l = item 1 CBD_Entrance or
        l = item 4 CBD_Entrance or l = item 5 CBD_Entrance or l = item 6
        ↪ CBD_Entrance
    [set-next-car-link l [end2] of l] #;; ask the vehicles to head south
    [set-next-car-link l [end1] of l] #;; ask the vehicles to head north
]

create-cars round (item ticks v_soparo * car_ratio / 60) [
 car-info
 let l item 7 CBD_Entrance
 ifelse l = item 0 CBD_Entrance or l = item 1 CBD_Entrance or
        l = item 4 CBD_Entrance or l = item 5 CBD_Entrance or l = item 6
        ↪ CBD_Entrance
    [set-next-car-link l [end2] of l] #;; ask the vehicles to head south
    [set-next-car-link l [end1] of l] #;; ask the vehicles to head north
]


create-cars round (item ticks v_saemunan * car_ratio / 60) [
 car-info
 let l item 8 CBD_Entrance
 ifelse l = item 0 CBD_Entrance or l = item 1 CBD_Entrance or
```

```
          l = item 4 CBD_Entrance or l = item 5 CBD_Entrance or l = item 6
          ↪ CBD_Entrance
      [set-next-car-link l [end2] of l] #;; ask the vehicles to head south
      [set-next-car-link l [end1] of l] #;; ask the vehicles to head north
  ]

  create-cars round (item ticks v_ssm * car_ratio / 60) [
   car-info
   let l item 9 CBD_Entrance
   ifelse l = item 0 CBD_Entrance or l = item 1 CBD_Entrance or
          l = item 4 CBD_Entrance or l = item 5 CBD_Entrance or l = item 6
          ↪ CBD_Entrance
      [set-next-car-link l [end2] of l] #;; ask the vehicles to head south
      [set-next-car-link l [end1] of l] #;; ask the vehicles to head north
  ]
end

to car-info
   set random-car true
   set parked false
   set size 2
   set speed random-float 2
   set unwell_history false
   let f_type# random 3
   if f_type# = 0 [set fueltype "Gasoline"]
   if f_type# = 1 [set fueltype "Diesel"]
   if f_type# = 2 [set fueltype "LPG"]
end


to drive-out-of-cbd
  let is-weekend? item 6 table:get pm10-road (ticks + 1)
  let what-time?  item 1 table:get pm10-road (ticks + 1)


  ask cars with [random-car] [ if any? cars-on nodes-here with [endpoint?][die]]
  ask cars with [not random-car] [ if is-weekend? and what-time? >= (8 + random 2)
  ↪  and what-time? < 22 and any? cars-on nodes-here with [endpoint?][set speed 0]
    if is-weekend? and what-time? < 8 or what-time? >= 22 and any? cars-on
    ↪  nodes-here with [endpoint?][move-to origin]
  ]
end
```

```
to kill-cars
  let night item 1 table:get pm10-back (ticks + 1)

  if (night >= 8 and night < 23 )
   [ ask n-of (int(.05 * car_ratio * count cars with [random-car = true])) cars
   ↪  with [random-car = true] [die]]

  if (night >= 23 and night < 24 or night < 4)
   [ ask n-of (int(.02 * car_ratio * count cars with [random-car = true])) cars
   ↪  with [random-car = true] [die]]
end


to move-drivers
    ask drivers with [my_car != nobody][
    if ([parked] of my_car = true) and ([current] of my_car = [goal] of my_car) [
    ↪  move-to [destination] of my_car ]
    if ([parked] of my_car = true) and ([current] of my_car = [origin] of my_car) [
    ↪  move-to [origin] of my_car ]
  ]
end

to add-employees
  let hours item 1 table:get pm10-back (ticks + 1)
  let mins item 4 table:get pm10-back (ticks + 1)
  let is-weekend? item 6 table:get pm10-road (ticks + 1)

  ask employees [
    if (hours = leave-home-hour and mins = leave-home-mins) and is-weekend? = false
    ↪  [
      move-to origin
      set Heuristic distance goal
      set current patch-here
      set direction 1
      set arrive-tick ticks
      set unwell_history false
    ]
  ]
end
```

```
to move-employees
  ask employees with [goal != nobody][
    set current patch-here
    set Heuristic distance goal
    ifelse (patch-here != goal and direction = 1) [
      if (awareness = "no") [walk-to-work]
      if (awareness = "yes") [walk-with-awareness]
    ]
    [
      work-hard
      if (time-at-work <= 0) [ set direction -1 head-home ]
      ]
  ]
end

to walk-with-awareness
  set color 103
  set arrived? false
  set hidden? false
  set time-at-work 540 + random 61

  let choice1 neighbors with [pm10 < 100]
  let nearest1 min-one-of choice1 [distance [goal] of myself ]
  let choice2 min-n-of 3 neighbors [pm10]
  let nearest2 min-one-of choice2 [distance [goal] of myself ]

  ifelse nearest1 != nobody [move-to nearest1][move-to nearest2]
  face [goal] of self
  fd (0.7 + random-float .3)
 if [Heuristic] of self <= 1 [move-to goal]
end

to walk-to-work
  set color 103
  set arrived? false
  set hidden? false
  #;let choices neighbors with [pavement]
  let nearest min-one-of neighbors [distance [goal] of myself ]
  set time-at-work 540 + random 61
  face nearest
  fd (0.7 + random-float .3)
end
```

```
to head-home
  set Heuristic distance origin
  let hours item 1 table:get pm10-back (ticks + 1)
  let nearest min-one-of neighbors [distance [origin] of myself ]
  set color brown + 2
  set arrived? false
  set hidden? false
  ifelse current = myhome [set time-at-work 0 fd 0 set arrive-tick 0][ face nearest
  ↪  fd (.6 + random-float .4) ]
  if distance origin <= 1 [ move-to myhome ]
end

to work-hard
  let hours item 1 table:get pm10-back (ticks + 1)
  let minutes item 4 table:get pm10-back (ticks + 1)
  set arrived? true
  if (ticks - arrive-tick) > 80 [set hidden? true]
  set direction 0
  set color grey
  fd 0
  ifelse time-at-work < 0 [set direction -1] [set time-at-work time-at-work - 1 ]
end

#;;-----------------Set exposure & Impact--------------#;;
to pollute
#;Small Vehicles
  ask cars with [not parked and speed > 0][
    let polluting one-of [link-neighbors] of to-node
    let mycar self
    let pm10# (item 2 table:get pm10-back (ticks + 1)) + random-float (item 3
    ↪  table:get pm10-back (ticks + 1))

    ask patches in-cone 2.5 90 [
      set pcolor grey + 2
      set pm10 pm10# + [total-emission] of mycar  #;; Removing 23\% the PM10 from
      ↪  vehicle contribution
                                                  #;; According to Weinbruch et al
                                                  ↪  (2014)
    ]
  ]
```

```
  ask patches with [pcolor != (grey + 2) and is-research-area? = true][
    set pm10 (item 2 table:get pm10-back (ticks + 1)) + random-float (item 3
    ↪  table:get pm10-back (ticks + 1))
  ]

  #;; cars
  ask cars with [owner != 0][
    ask origin [set pm10_indoor ([pm10] of patch-here) * (.2 + random-float .5 ) ]
  ]

  #;; set buildings
  ask buildings [
   ask patch-here [ set pm10_indoor pm10 * (.2 + random-float .5 )  ]
  ]

  #;; in the left bottom corner
  ask patch max-pxcor max-pycor
    [ let night item 1 table:get pm10-others (ticks + 1)
      ifelse night > 23 or night < 6
      [ set pm10 ((item 2 table:get pm10-others (ticks + 1)) +
        random-float (item 3 table:get pm10-others (ticks + 1))) * .25  ]
      [ set pm10 (item 2 table:get pm10-others (ticks + 1)) +
        random-float (item 3 table:get pm10-others (ticks + 1)) * (1 - random-float
        ↪  .75 )]
    ]

  #;; for resident vehicles going outside
  ask patches with [is-endpoint?]
    [ let night item 1 table:get pm10-others (ticks + 1)
      ifelse night > 23 or night < 6
      [ set pm10 ((item 2 table:get pm10-others (ticks + 1)) +
        random-float (item 3 table:get pm10-others (ticks + 1))) * (.25 +
        ↪  random-float .5)  ]
      [ set pm10 (item 2 table:get pm10-others (ticks + 1)) +
        random-float (item 3 table:get pm10-others (ticks + 1)) ]
    ]
end

to fadeout
  ask patches with [pcolor = (grey + 2)][
    ifelse countdown <= 0
      [ set pcolor white
```

```
        set countdown random 3 ]
      [ set countdown countdown - 1 ]
  ]
end

#;;---------------- Health Loss & Recover ------------------#;;
to health-loss
  ask employees [
    if ([pm10] of patch-here >= 100 and arrived? = false)
      [set health health - ((random-float health_loss) * (310 - health))]
    if ([pm10_indoor] of patch-here != nobody and [pm10_indoor] of patch-here >=
    ↪  100 and arrived? = true)
      [set health health - ((random-float health_loss) * (310 - health))]
    if health < 100 [set unwell_history true ]
    if health < 0 [set health 0]
  ]

  ask cars with [owner != 0][
    if not parked and ([pm10] of patch-here * 0.7) >= 100 [set health health -
    ↪  ((random-float health_loss) * (310 - health)) ]
    if parked and [pm10_indoor] of destination >= 100 [set health health -
    ↪  ((random-float health_loss) * (310 - health)) ]
    if health < 100 [set unwell_history true ]
    if health < 0 [set health 0]
  ]

  ask drivers with [my_car != nobody][
    if [parked] of my_car = false [ set health [health] of my_car ]]
end

to health-recovery
  ask employees with [current = myhome or current = goal and health < 250]
     [ if health >= 0 and health < 100 [set health health + (medication +
     ↪  random-float medication) ]]
  ask cars with [owner != 0 and parked and health >= 0 and health < 250]
     [ if health >= 0 and health < 100 [set health health + (medication +
     ↪  random-float medication) ]]
  ask drivers with [my_car != nobody][ set health [health] of my_car]
end

to validation-plot
  set-current-plot "Validation"
```

```
  set-current-plot-pen "Jongno" plot([pm10] of patch 134 98)
  set-current-plot-pen "Jung" plot([pm10] of patch 38 68)
  set-current-plot-pen "Jongno Kerb" plot([pm10] of patch 105 95) #;; Jongno kerb
  set-current-plot-pen "Seoul Stn" plot([pm10] of patch 40 42)
end

to age-plot
 set-current-plot "Health by Age"
 set-current-plot-pen "Drivers" plot(count(drivers with [health < 100]) / count
 ↪  drivers * 100)
 set-current-plot-pen "Walkers" plot(count(employees with [health < 100]) / count
 ↪  employees * 100)

end
```

Appendix C

Chapter 3: Outcomes

C.1 Universal Kriging

C.1.1 NO_2: Semivariogram

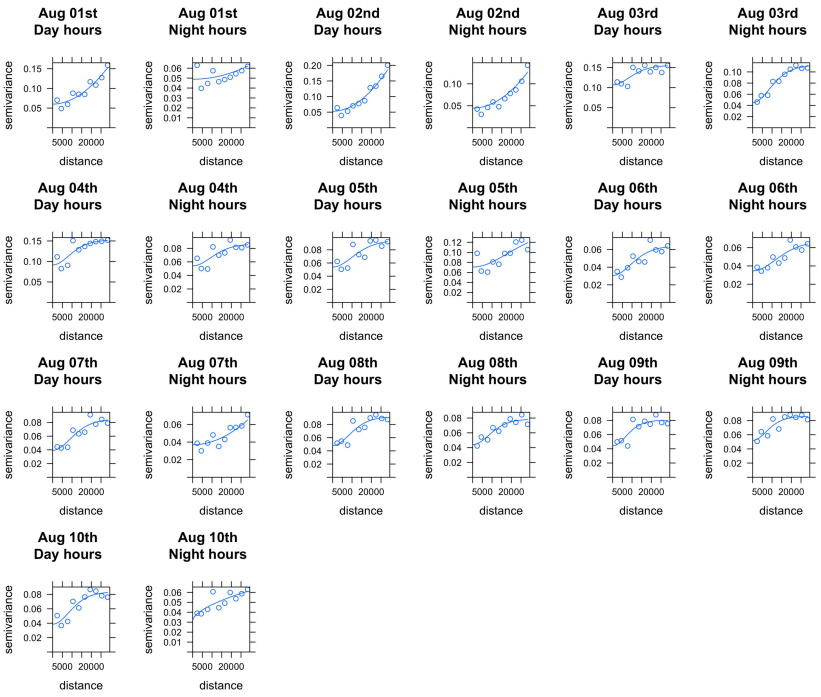


Fig. C.1 NO_2: Semivariogram for Kriged Maps between Aug 1st-10th

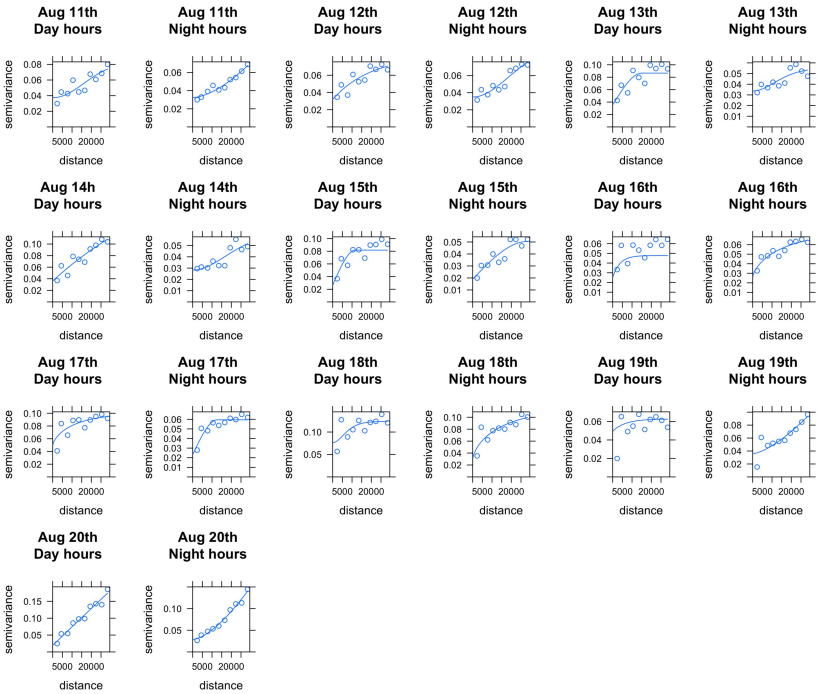


Fig. C.2 NO_2: Semivariogram for Kriged Maps between Aug 11th-20th

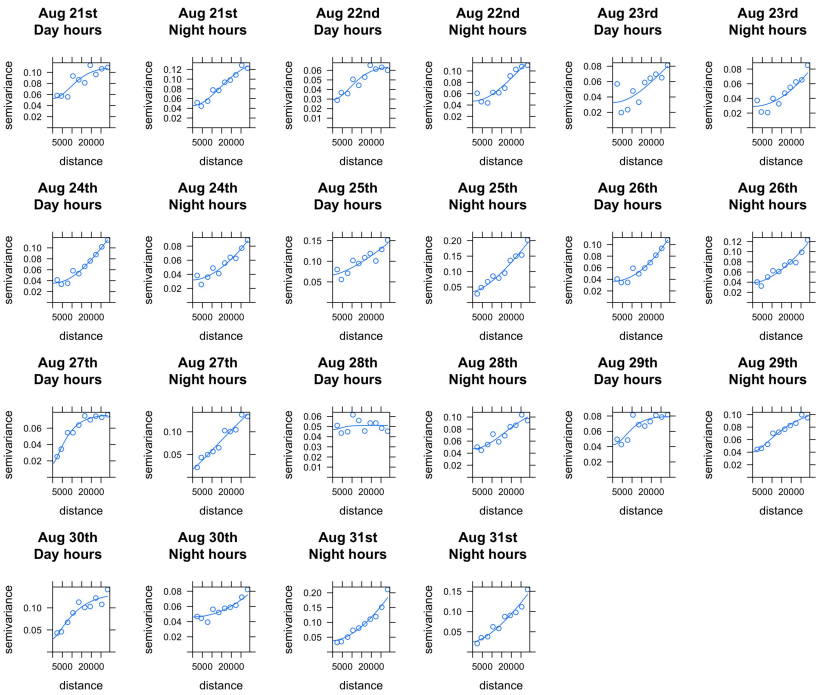


Fig. C.3 NO_2: Semivariogram for Kriged Maps between Aug 21th-31st

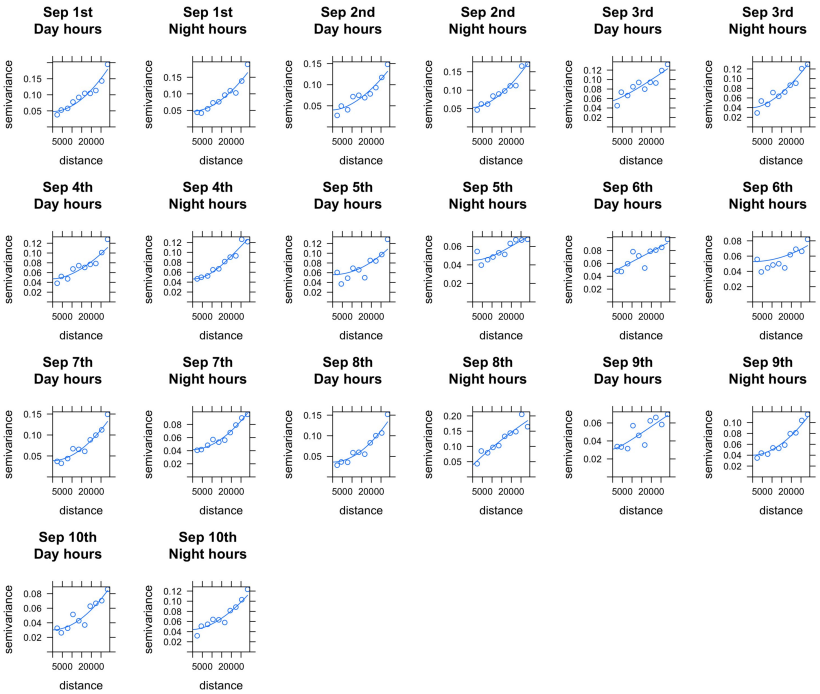


Fig. C.4 NO_2: Semivariogram for Kriged Maps between Sep 1st-10th

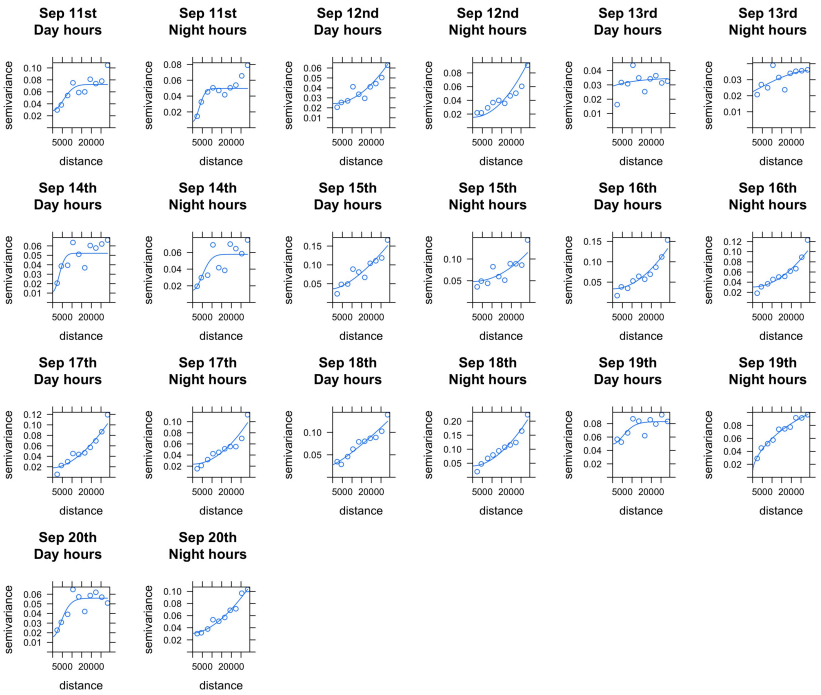


Fig. C.5 NO_2: Semivariogram for Kriged Maps between Sep 11th-20th

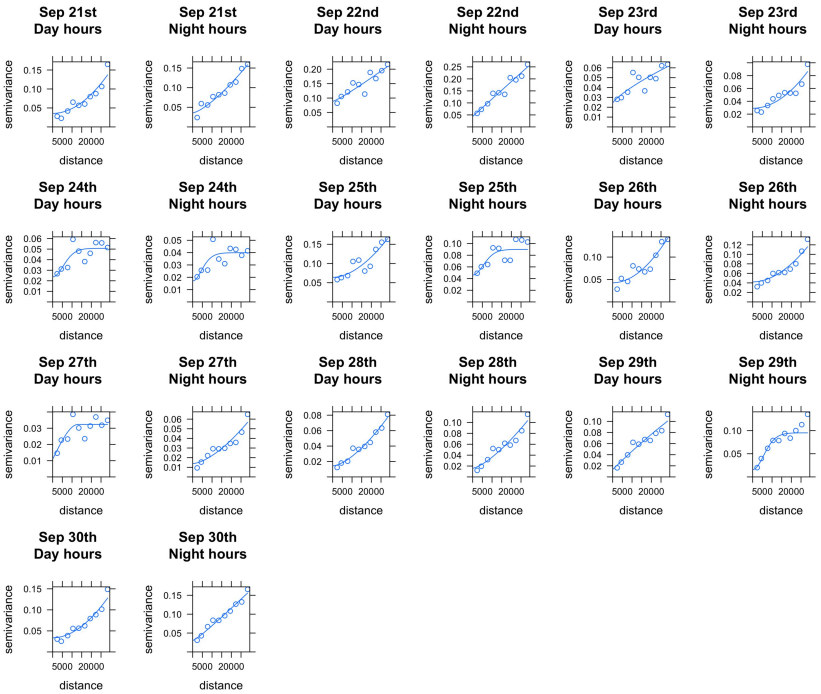


Fig. C.6 NO_2: Semivariogram for Kriged Maps between Sep 21st-30th

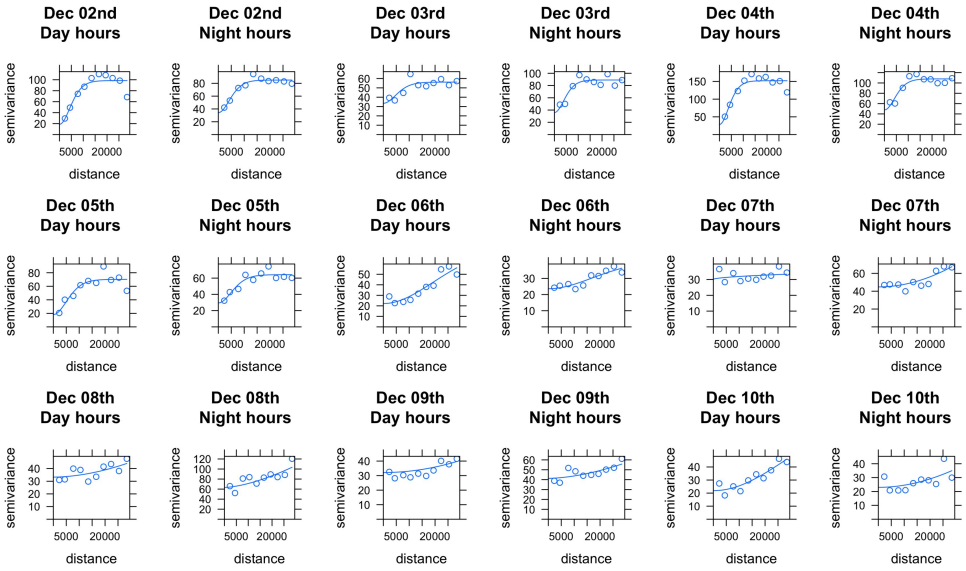


Fig. C.7 NO_2: Semivariogram for Kriged Maps between Dec 1st-10th

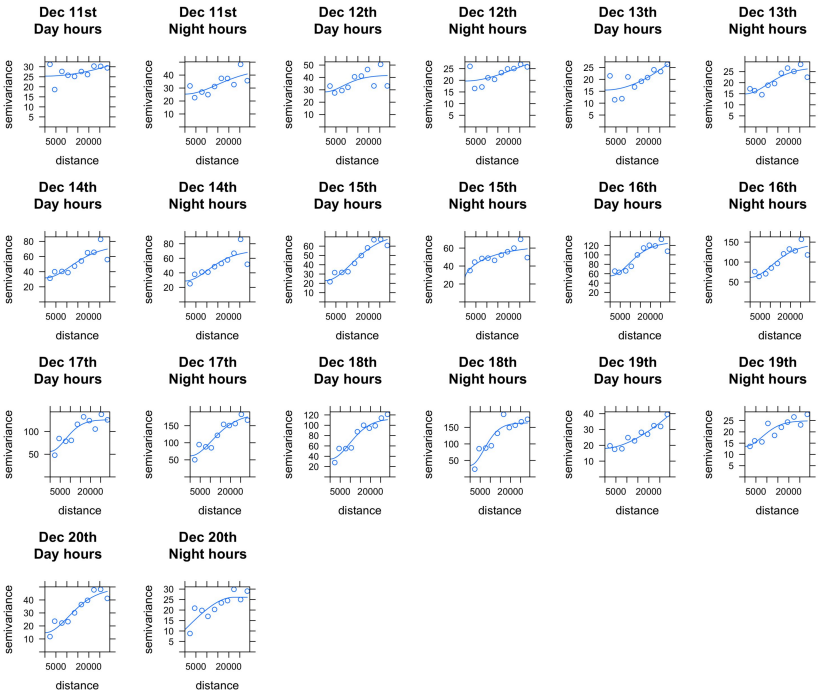


Fig. C.8 NO_2: Semivariogram for Kriged Maps between Dec 11th-20th

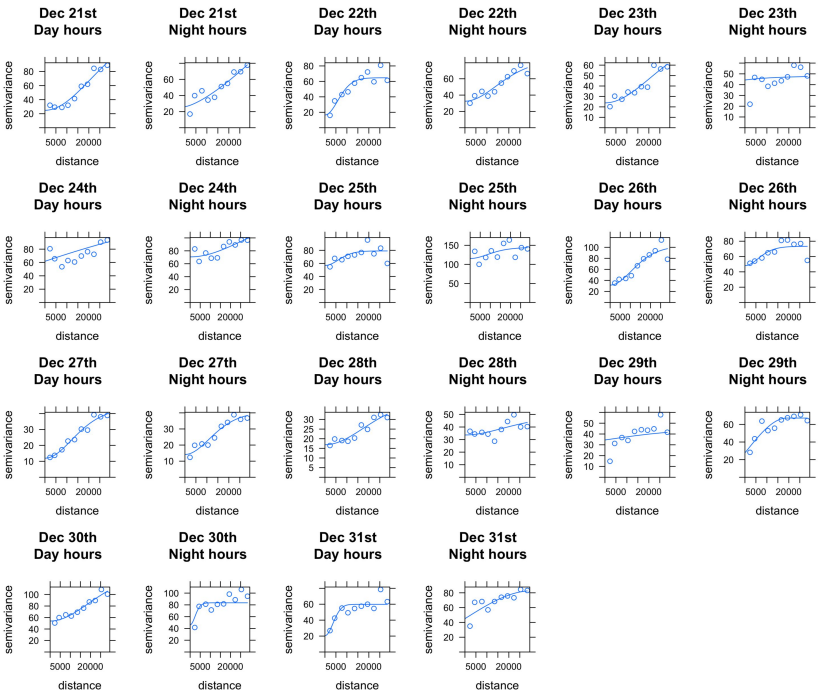


Fig. C.9 NO_2: Semivariogram for Kriged Maps between Dec 21th-31st

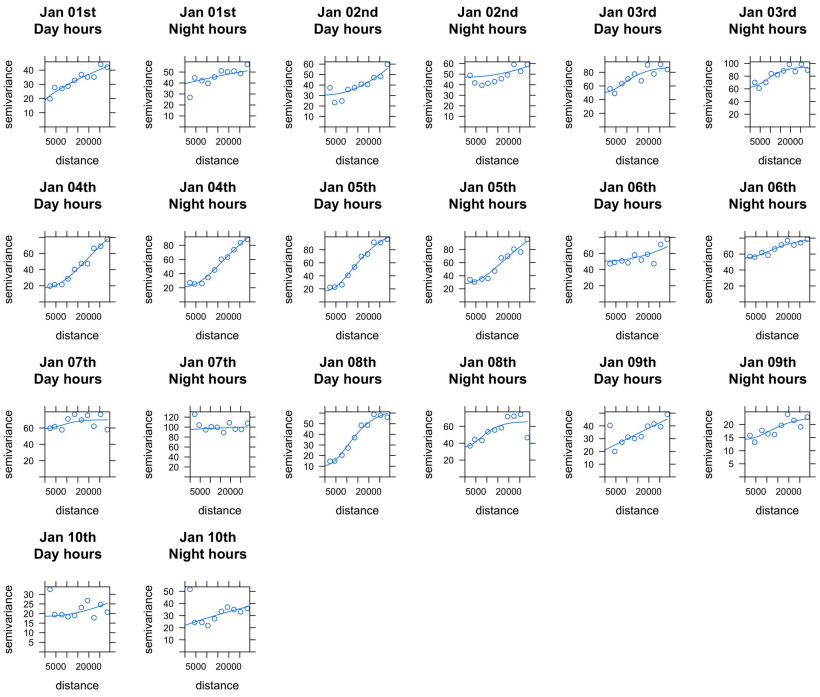


Fig. C.10 NO_2: Semivariogram for Kriged Maps between Jan 1st-10th

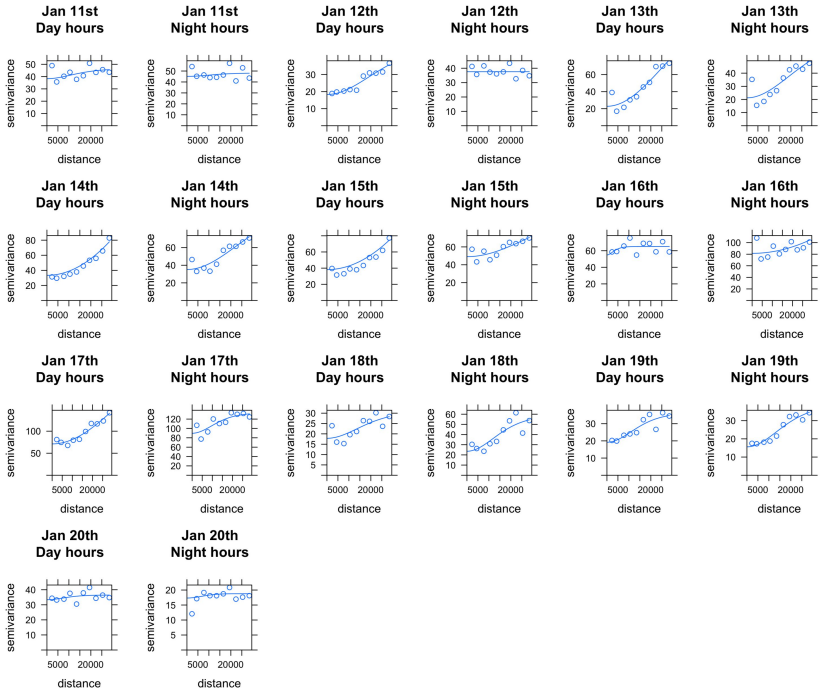


Fig. C.11 NO_2: Semivariogram for Kriged Maps between Jan 11th-20th

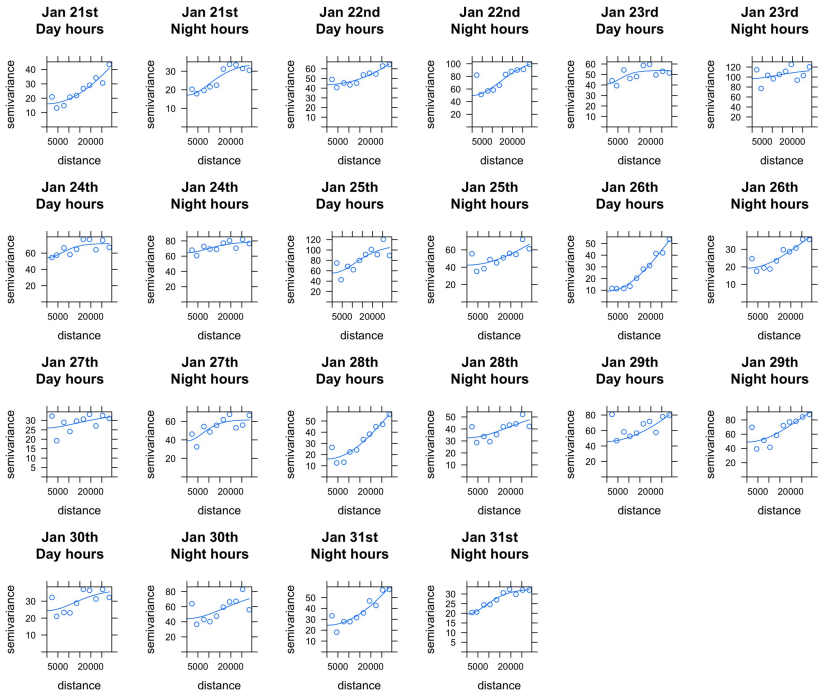


Fig. C.12 NO_2: Semivariogram for Kriged Maps between Jan 21th-31st

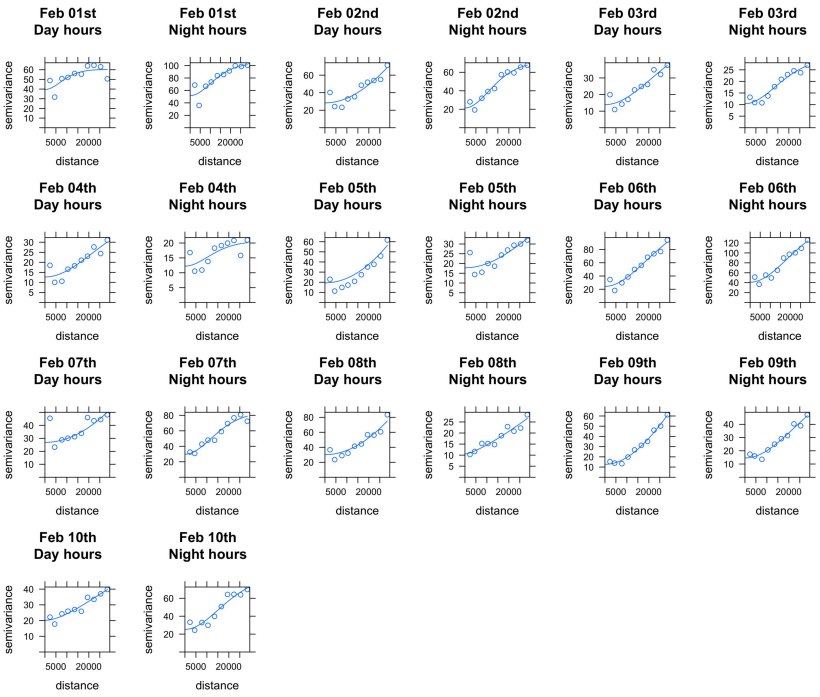


Fig. C.13 NO_2: Semivariogram for Kriged Maps between Feb 1st-10th

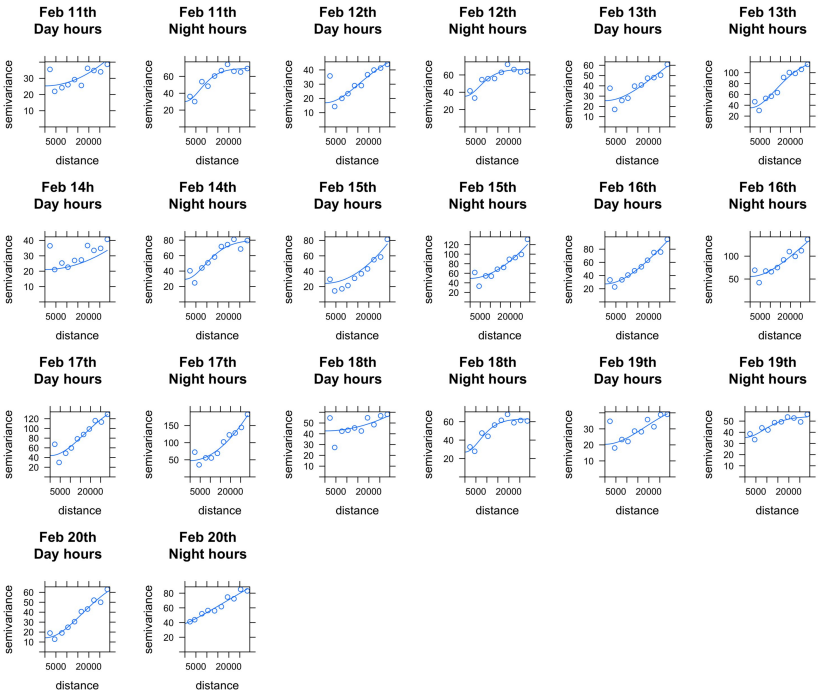


Fig. C.14 NO_2: Semivariogram for Kriged Maps between Feb 11th-20th

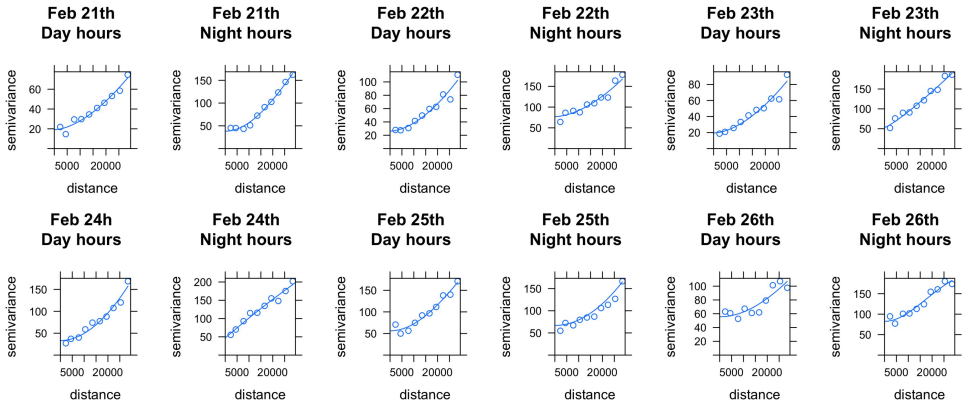


Fig. C.15 NO_2: Semivariogram for Kriged Maps between Feb 21th-28th

C.1.2 NO_2: Kriging - Predictions

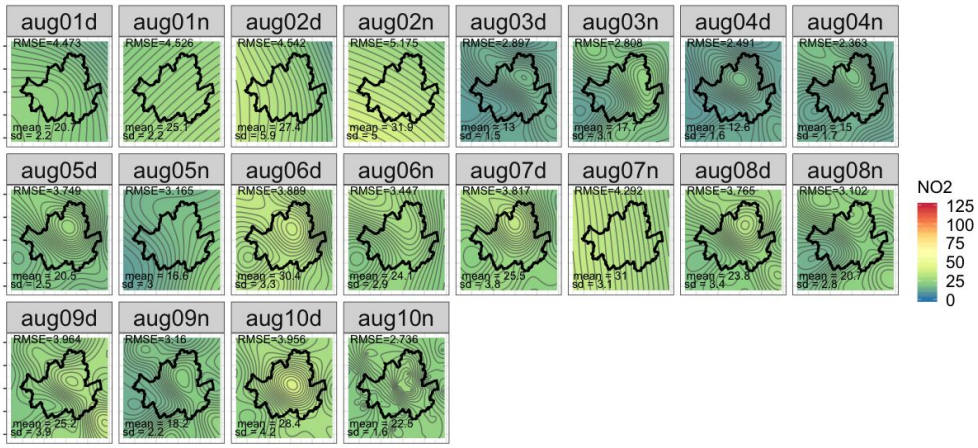


Fig. C.16 NO_2: Kriged Maps between Aug 1st-10th (units: ppb)

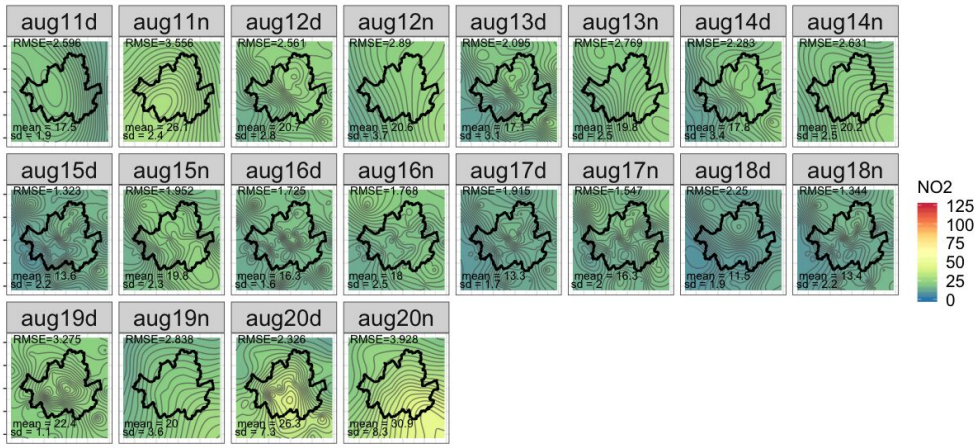


Fig. C.17 NO_2: Kriged Maps between Aug 11th-20th (units: ppb)

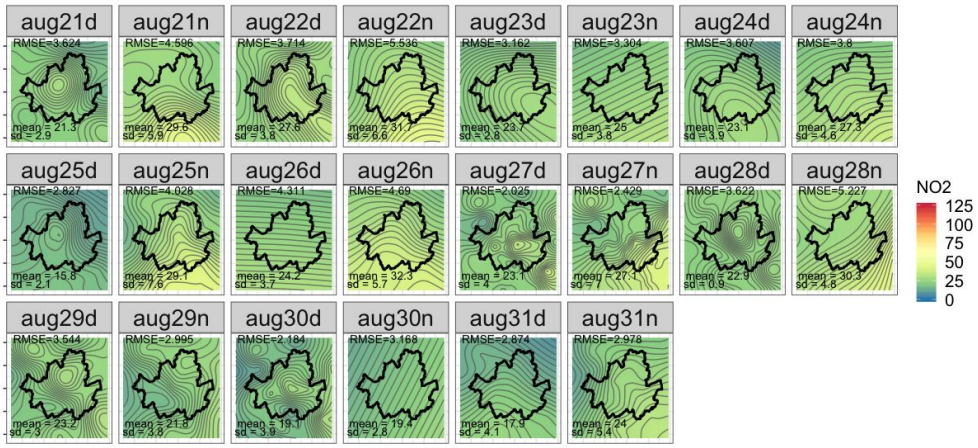


Fig. C.18 NO_2: Kriged Maps between Aug 21th-31st (units: ppb)

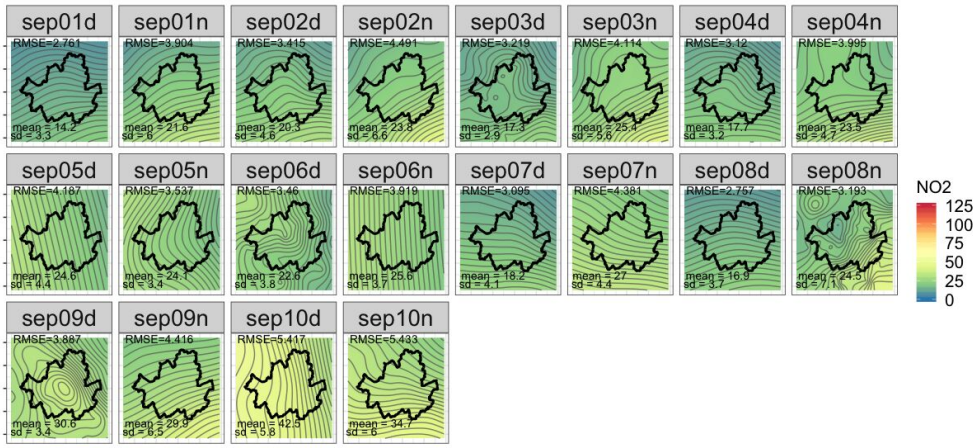


Fig. C.19 NO_2: Kriged Maps between Sep 1st-10th (units: ppb)

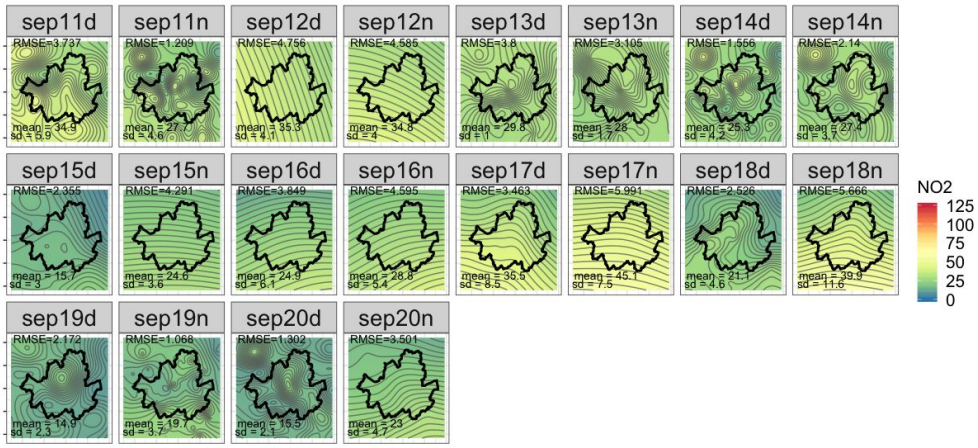


Fig. C.20 NO_2: Kriged Maps between Sep 11th-20th (units: ppb)

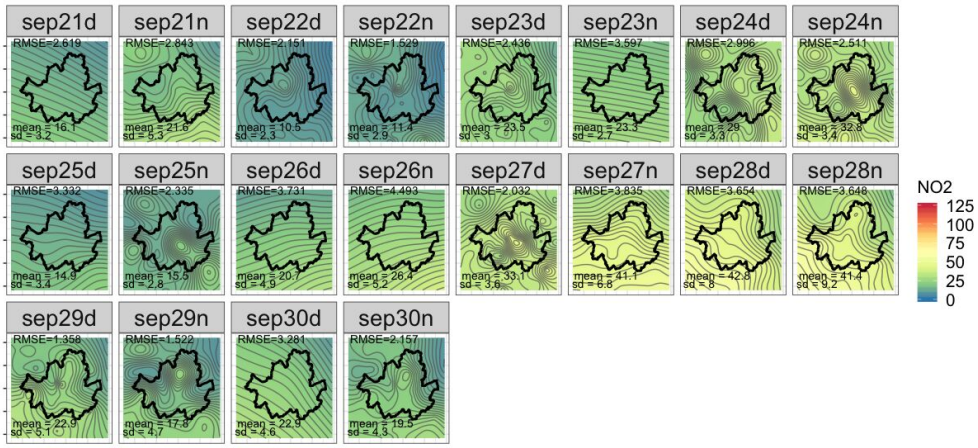


Fig. C.21 NO_2: Kriged Maps between Sep 21st-30th (units: ppb)

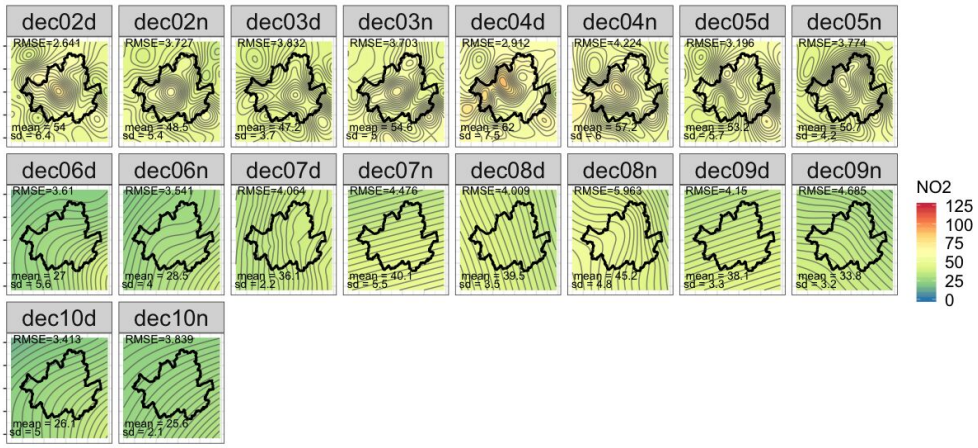


Fig. C.22 NO_2: Kriged Maps between Dec 1st-10th (units: ppb)

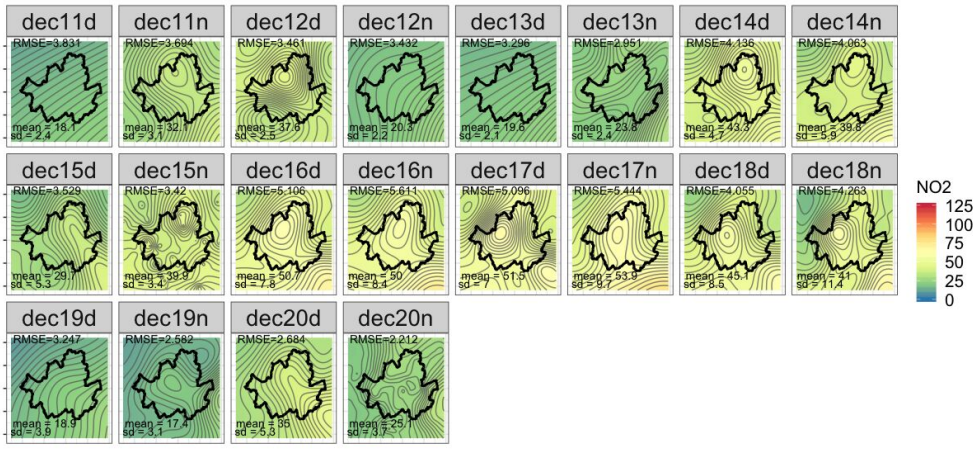


Fig. C.23 NO_2: Kriged Maps between Dec 11th-20th (units: ppb)

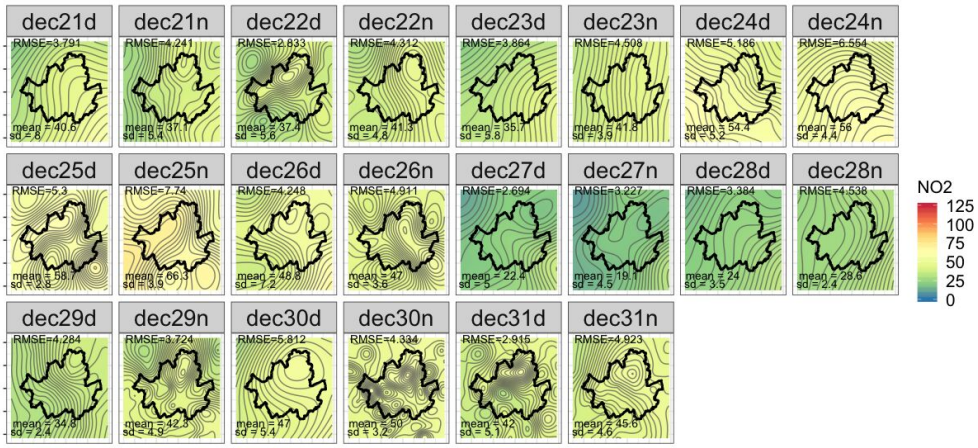


Fig. C.24 NO_2: Kriged Maps between Dec 21th-31st (units: ppb)

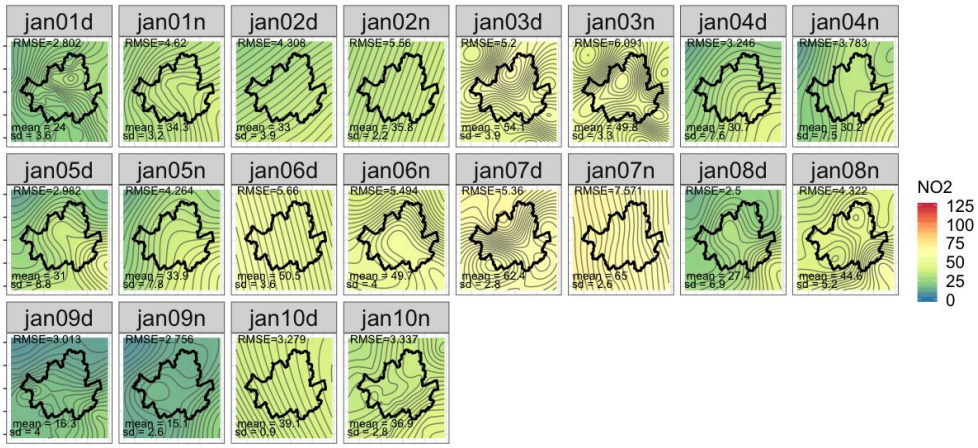


Fig. C.25 NO_2: Kriged Maps between Jan 1st-10th (units: ppb)

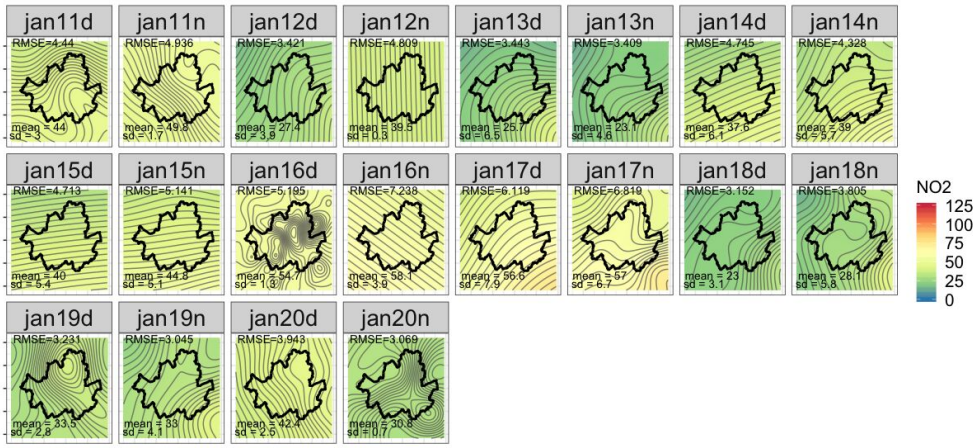


Fig. C.26 NO_2: Kriged Maps between Jan 11th-20th (units: ppb)

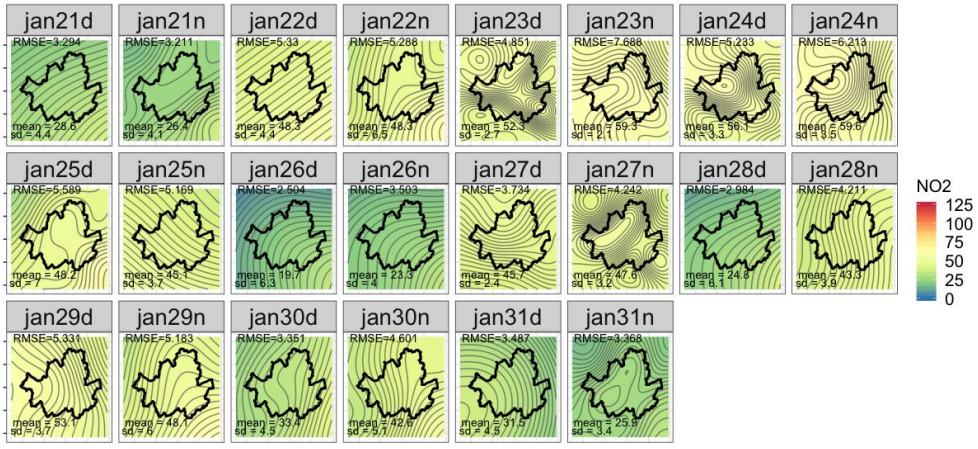


Fig. C.27 NO_2: Kriged Maps between Jan 21th-31st (units: ppb)

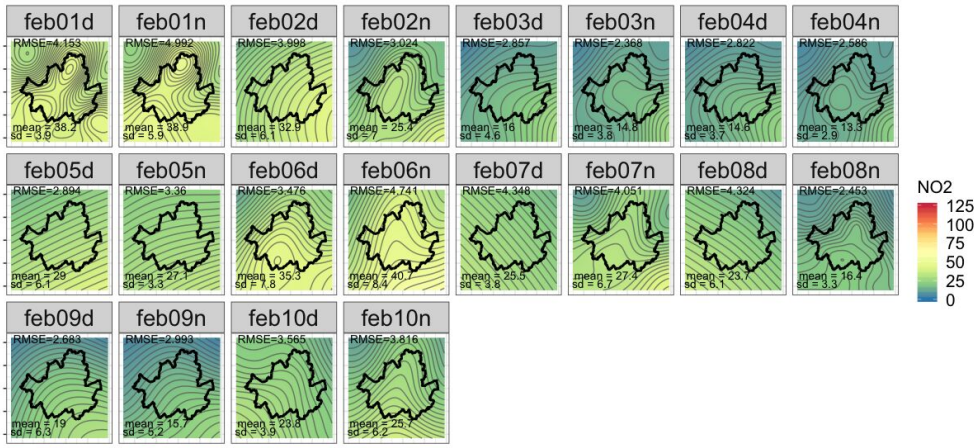


Fig. C.28 NO_2: Kriged Maps between Feb 1st-10th (units: ppb)

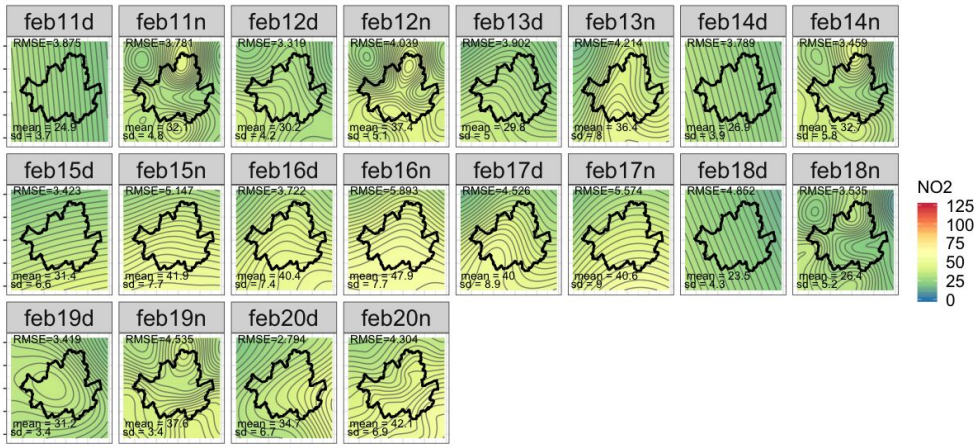


Fig. C.29 NO_2: Kriged Maps between Feb 11th-20th (units: ppb)

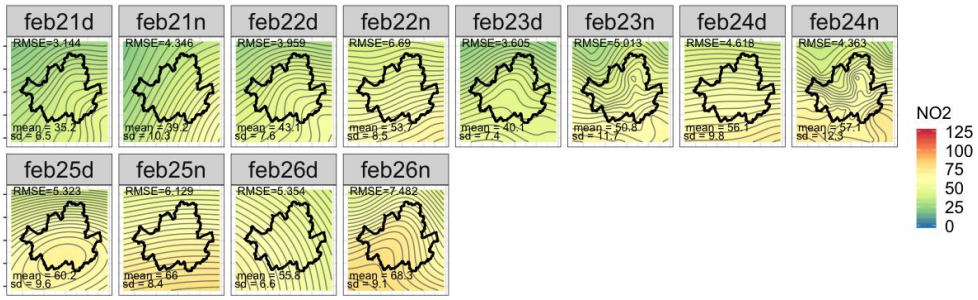


Fig. C.30 NO_2: Kriged Maps between Feb 21th-28th (units: ppb)

C.1.3 NO_2: Kriging - Predictions with Addtional Road Weight

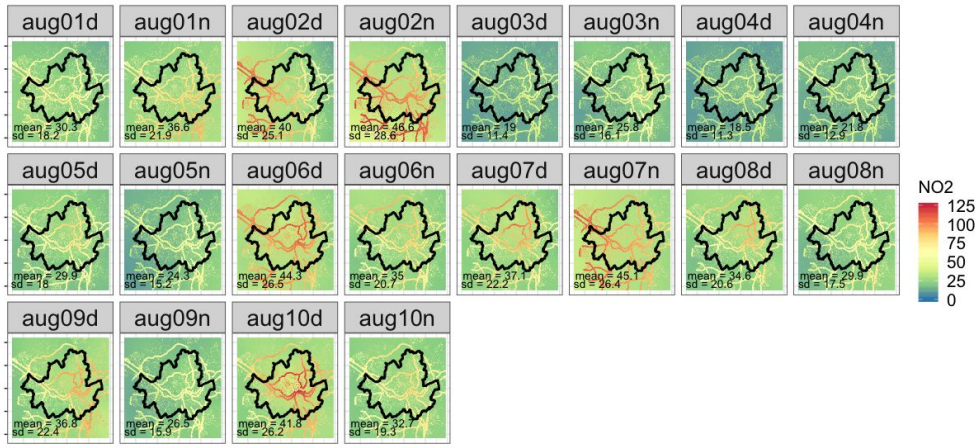


Fig. C.31 NO_2: Kriged Maps with Road Weights between Aug 1st-10th (units: ppb)

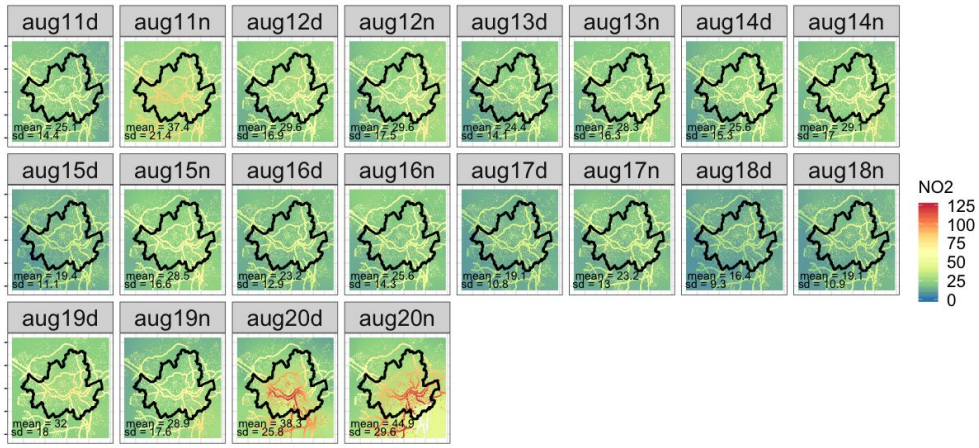


Fig. C.32 NO_2: Kriged Maps with Road Weights between Aug 11th-20th (units: ppb)

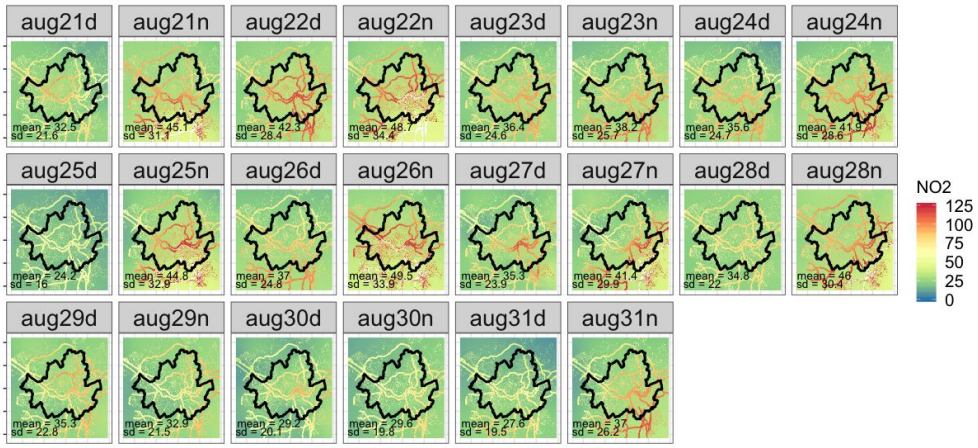


Fig. C.33 NO_2: Kriged Maps with Road Weights between Aug 21th-31st (units: ppb)

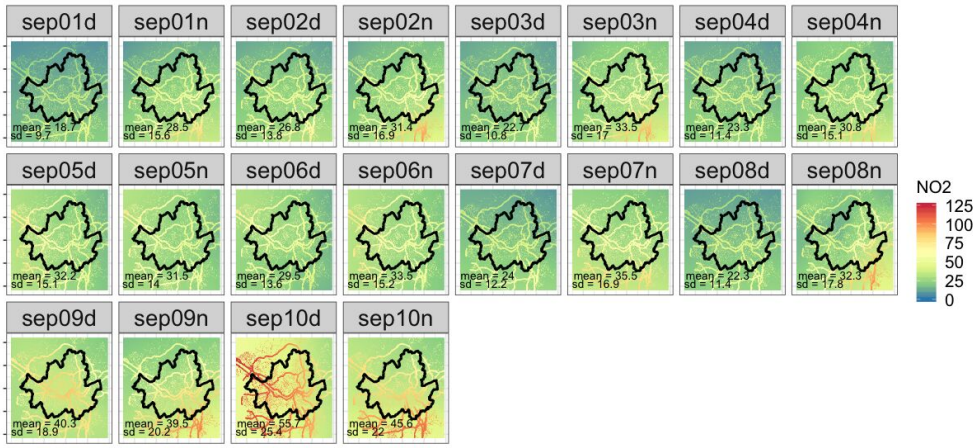


Fig. C.34 NO_2: Kriged Maps with Road Weights between Sep 1st-10th (units: ppb)

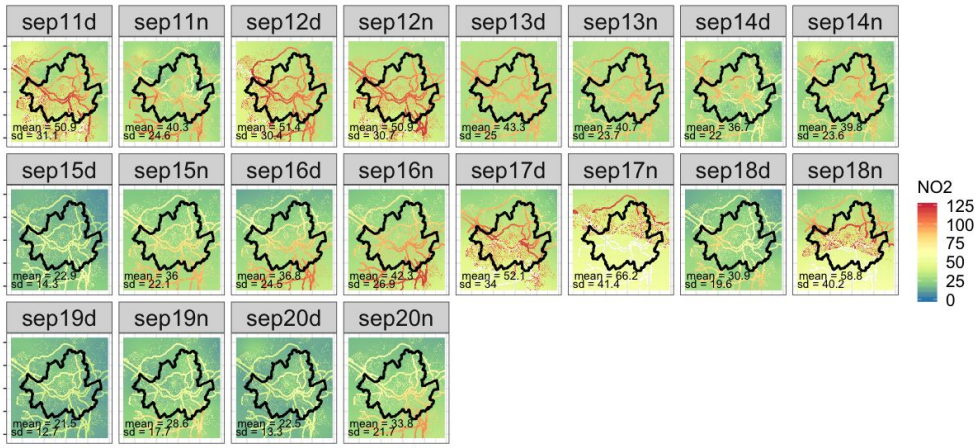


Fig. C.35 NO_2: Kriged Maps with Road Weights between Sep 11th-20th (units: ppb)

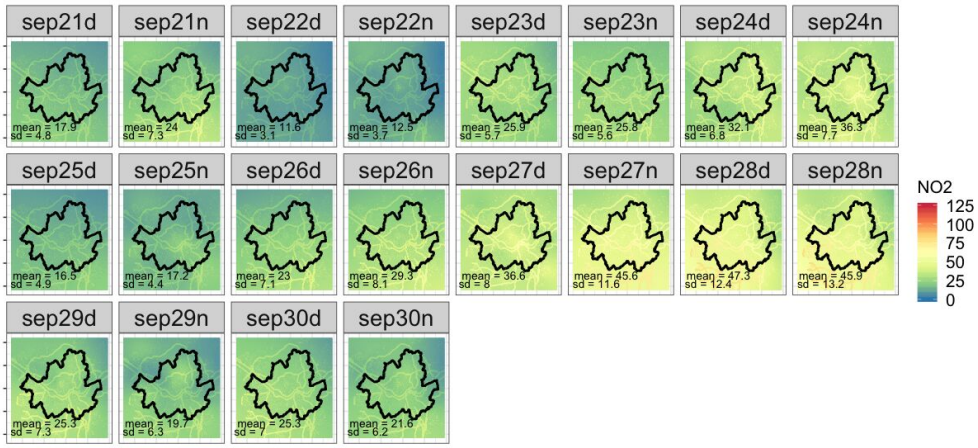


Fig. C.36 NO_2: Kriged Maps with Road Weights between Sep 21st-30th (units: ppb)

Fig. C.37 NO_2: Kriged Maps with Road Weights between Dec 1st-10th (units: ppb)

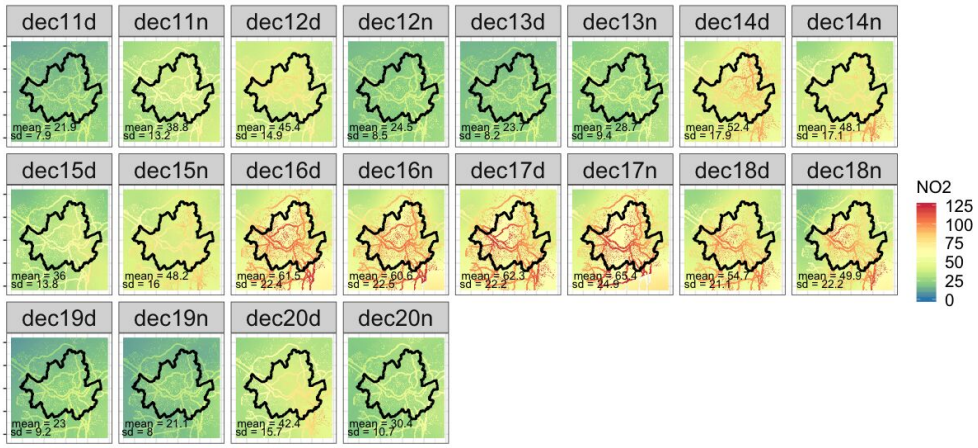


Fig. C.38 NO_2: Kriged Maps with Road Weights between Dec 11th-20th (units: ppb)

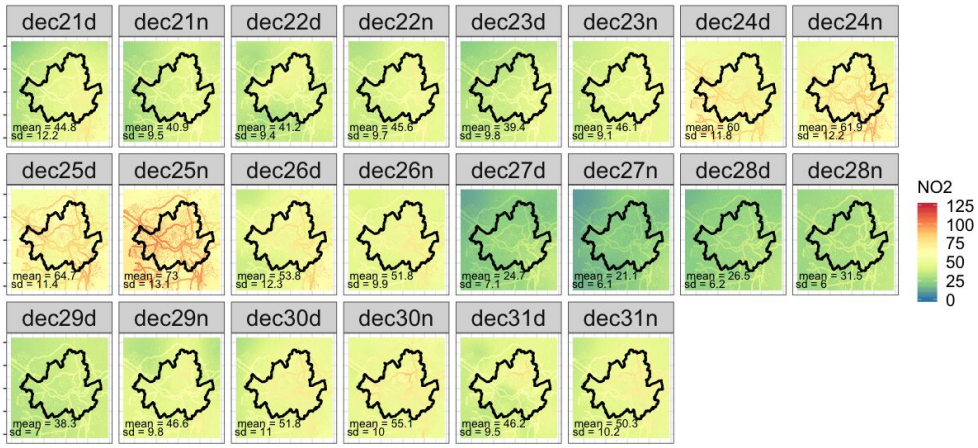


Fig. C.39 NO_2: Kriged Maps with Road Weights between Dec 21th-31st (units: ppb)

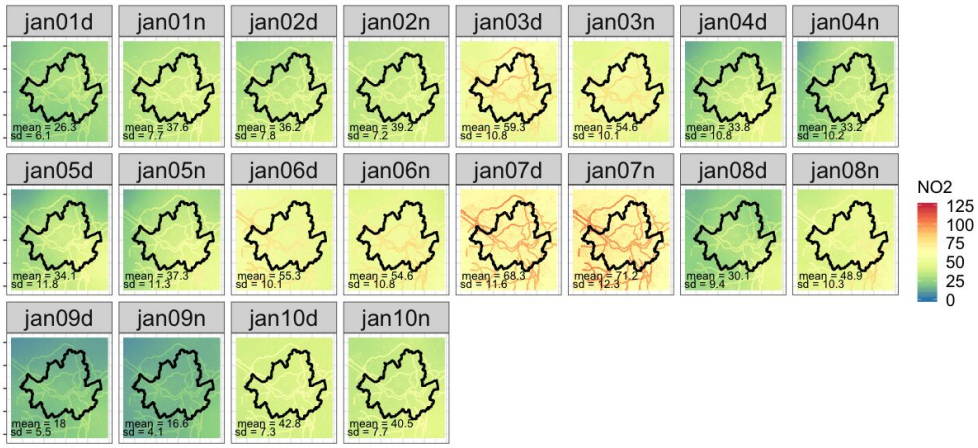


Fig. C.40 NO_2: Kriged Maps with Road Weights between Jan 1st-10th (units: ppb)

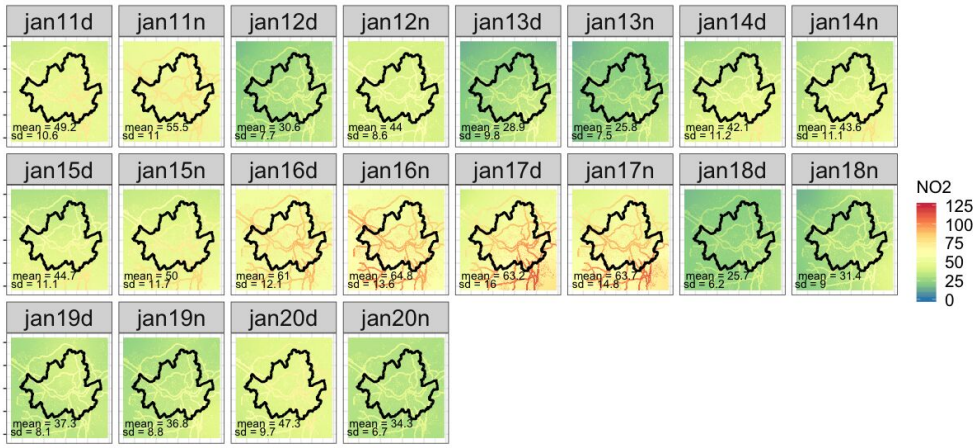


Fig. C.41 NO_2: Kriged Maps with Road Weights between Jan 11th-20th (units: ppb)

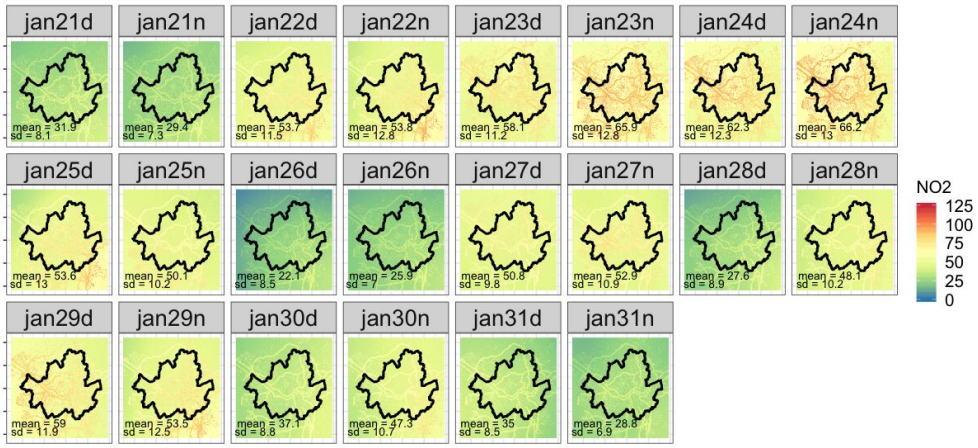


Fig. C.42 NO_2: Kriged Maps with Road Weights between Jan 21th-31st (units: ppb)

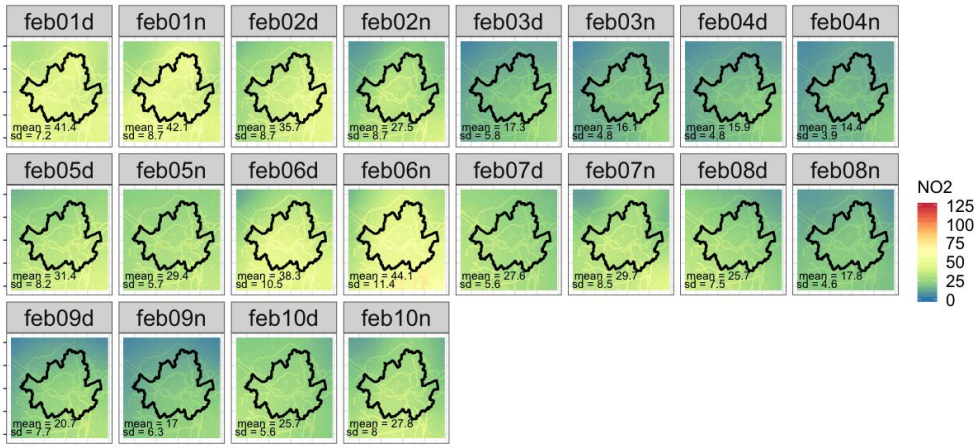


Fig. C.43 NO_2: Kriged Maps with Road Weights between Feb 1st-10th (units: ppb)

Fig. C.44 NO_2: Kriged Maps with Road Weights between Feb 11th-20th (units: ppb)

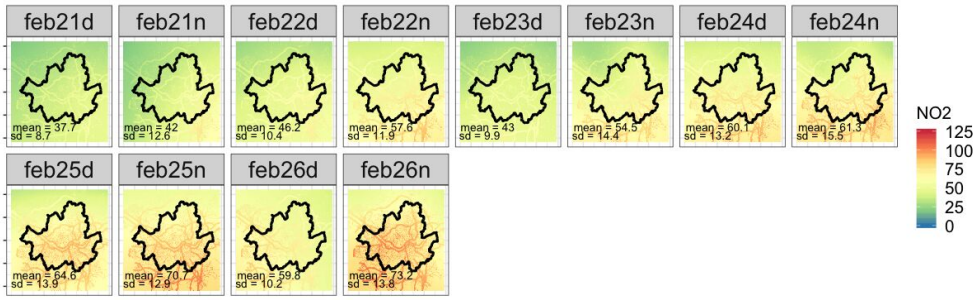


Fig. C.45 NO_2: Kriged Maps with Road Weights between Feb 21th-28th (units: ppb)

C.1.4 PM_{10}: Semivariogram

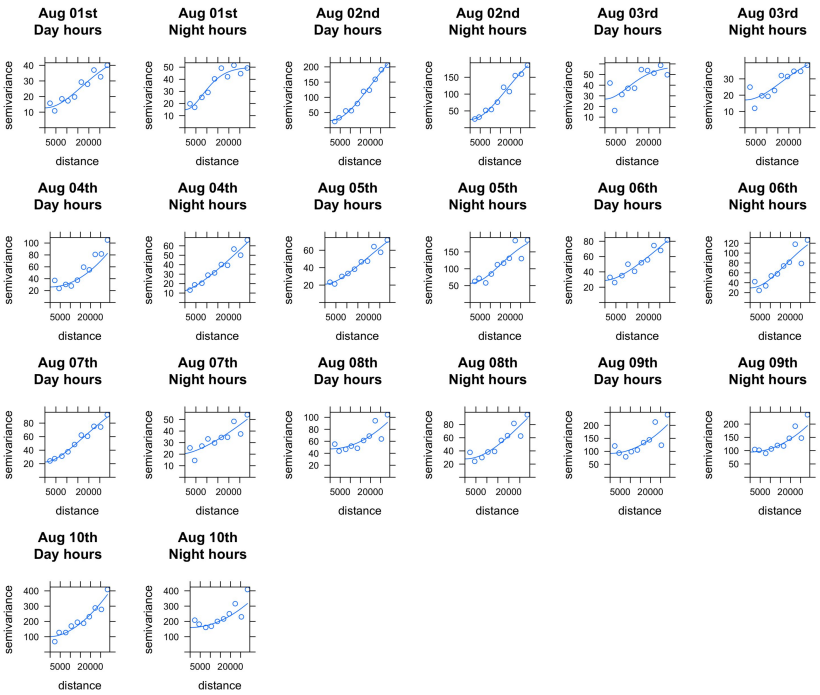


Fig. C.46 PM_{10}: Semivariogram for Kriged Maps between Aug 1st-10th

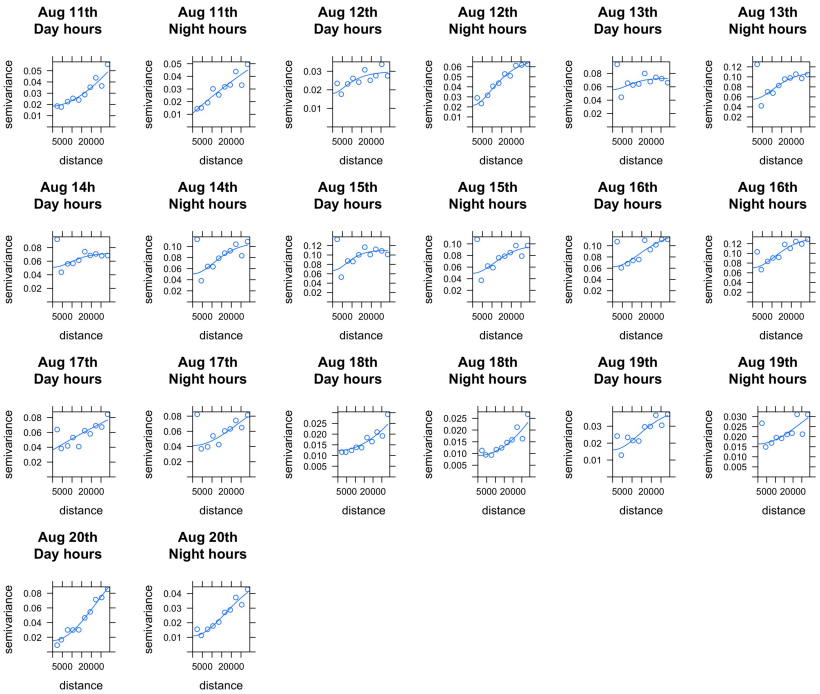


Fig. C.47 PM_{10}: Semivariogram for Kriged Maps between Aug 11th-20th

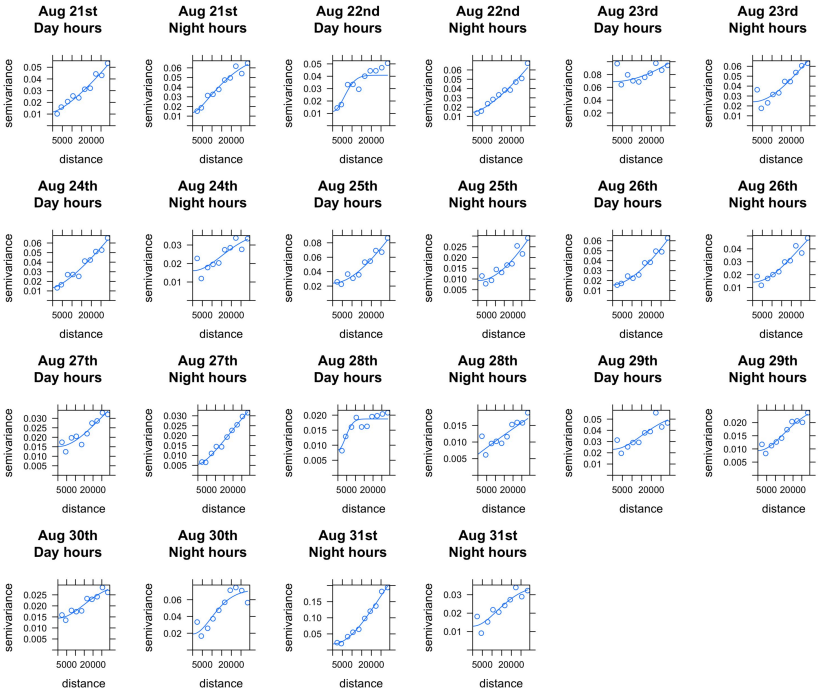


Fig. C.48 PM_{10}: Semivariogram for Kriged Maps between Aug 21th-31st

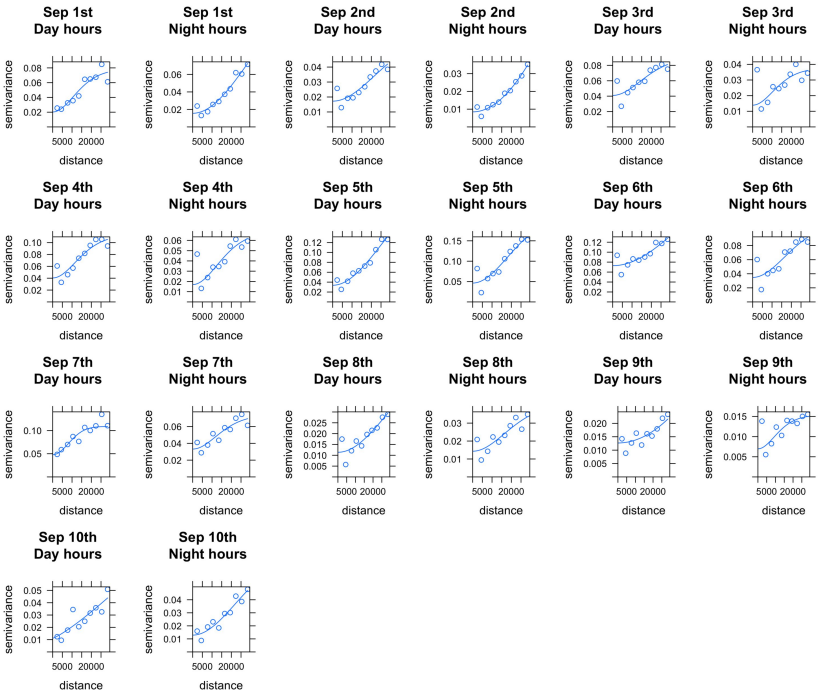


Fig. C.49 PM_{10}: Semivariogram for Kriged Maps between Sep 1st-10th

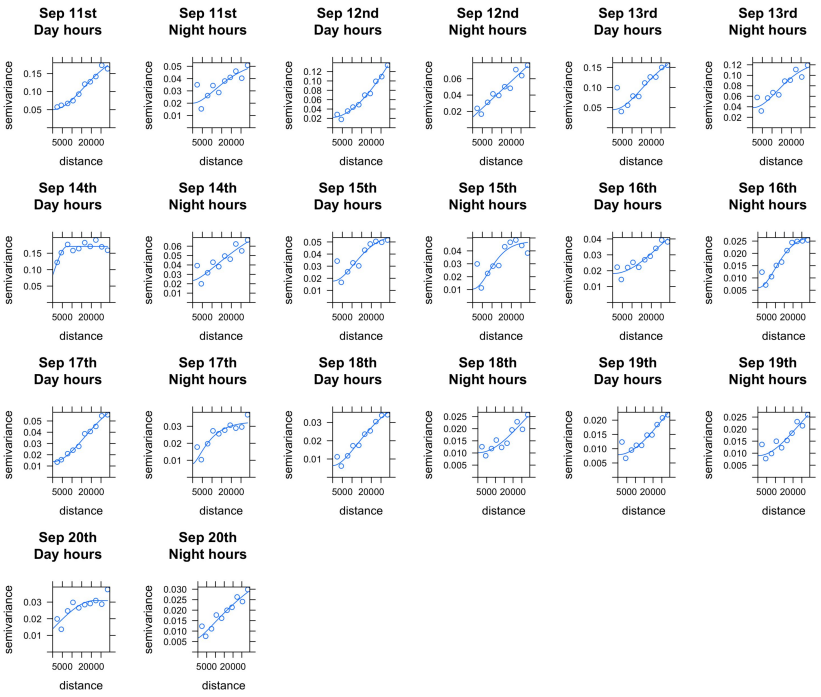


Fig. C.50 PM_{10}: Semivariogram for Kriged Maps between Sep 11th-20th

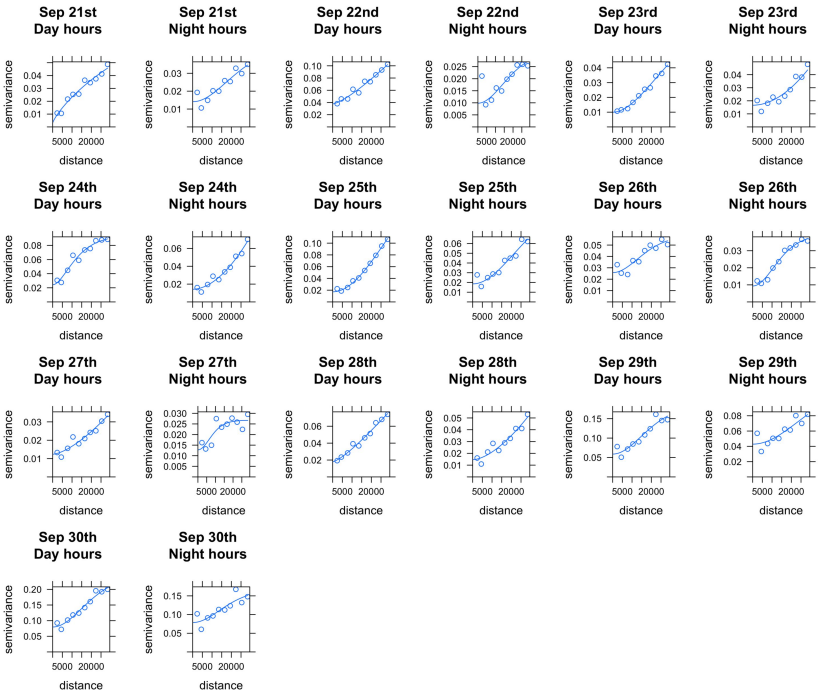


Fig. C.51 PM_{10}: Semivariogram for Kriged Maps between Sep 21st-30th

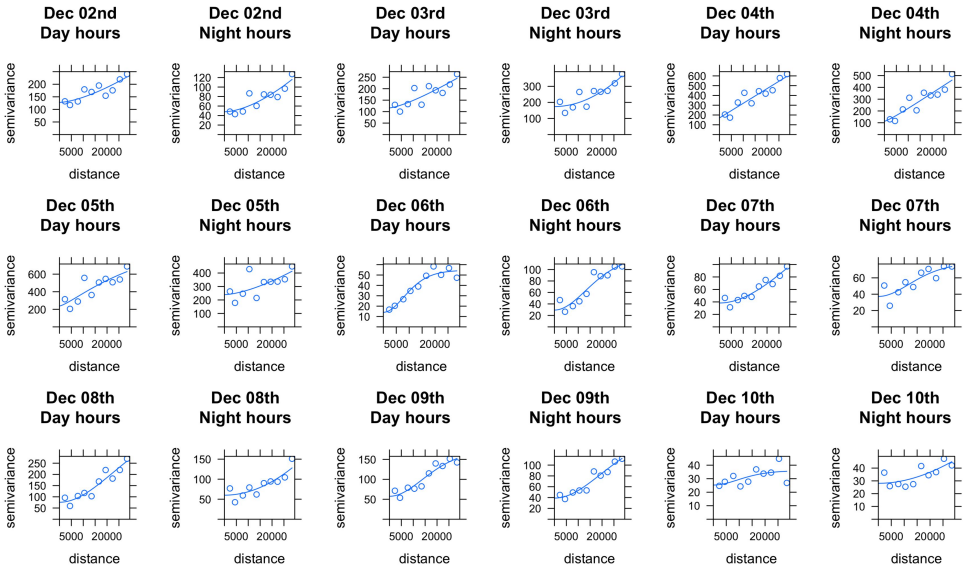


Fig. C.52 PM_{10}: Semivariogram for Kriged Maps between Dec 1st-10th

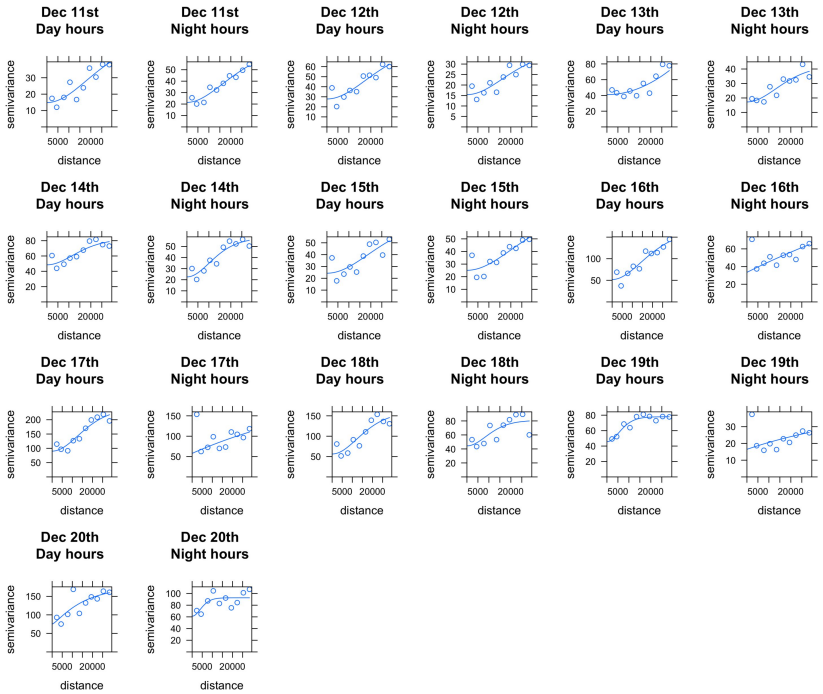


Fig. C.53 PM_{10}: Semivariogram for Kriged Maps between Dec 11th-20th

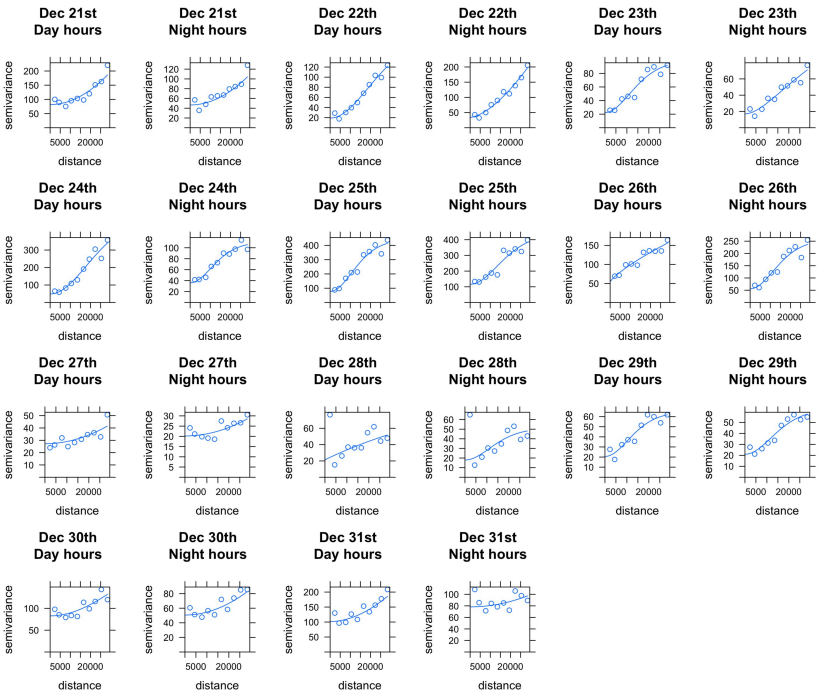


Fig. C.54 PM_{10}: Semivariogram for Kriged Maps between Dec 21th-31st

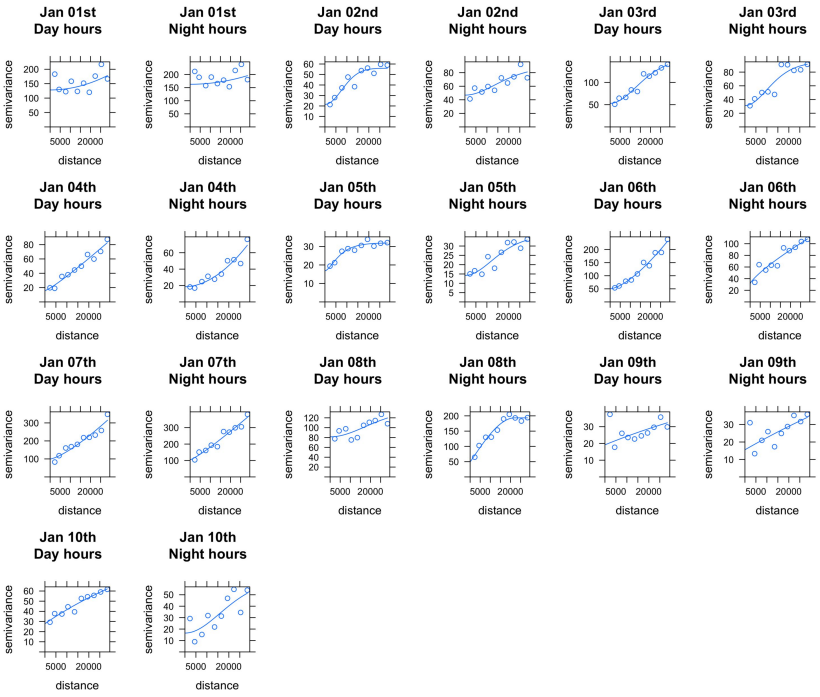


Fig. C.55 PM_{10}: Semivariogram for Kriged Maps between Jan 1st-10th

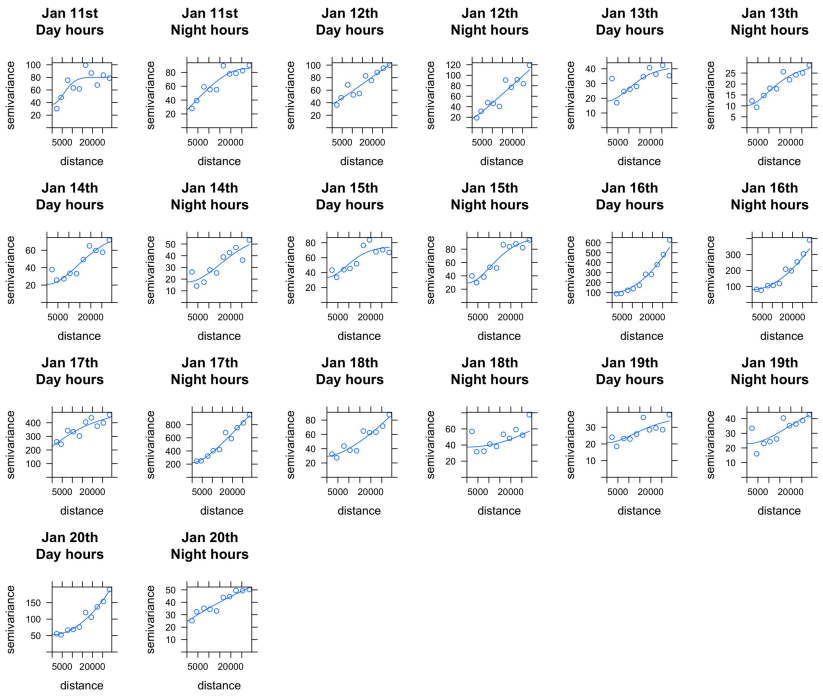


Fig. C.56 PM_{10}: Semivariogram for Kriged Maps between Jan 11th-20th

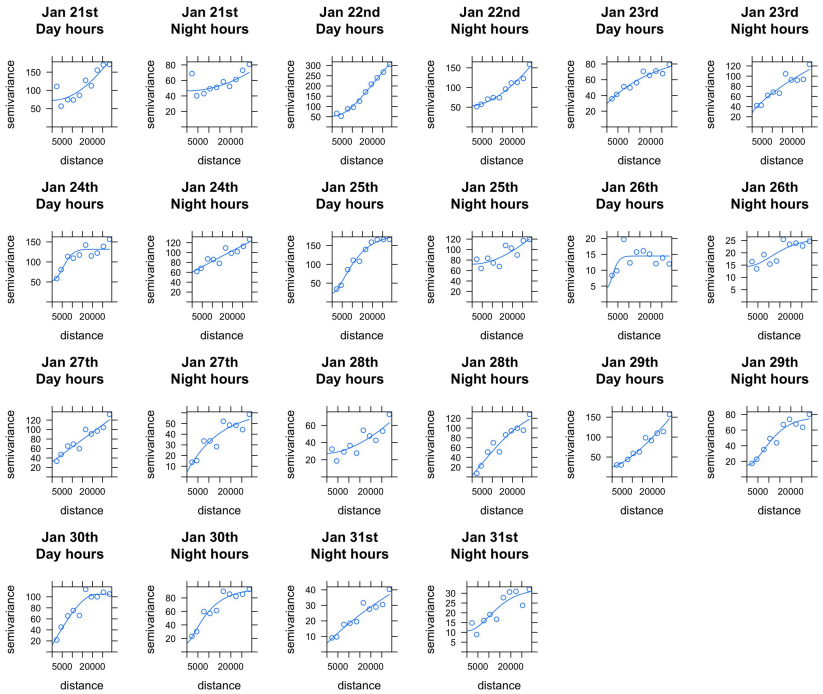


Fig. C.57 PM_{10}: Semivariogram for Kriged Maps between Jan 21th-31st

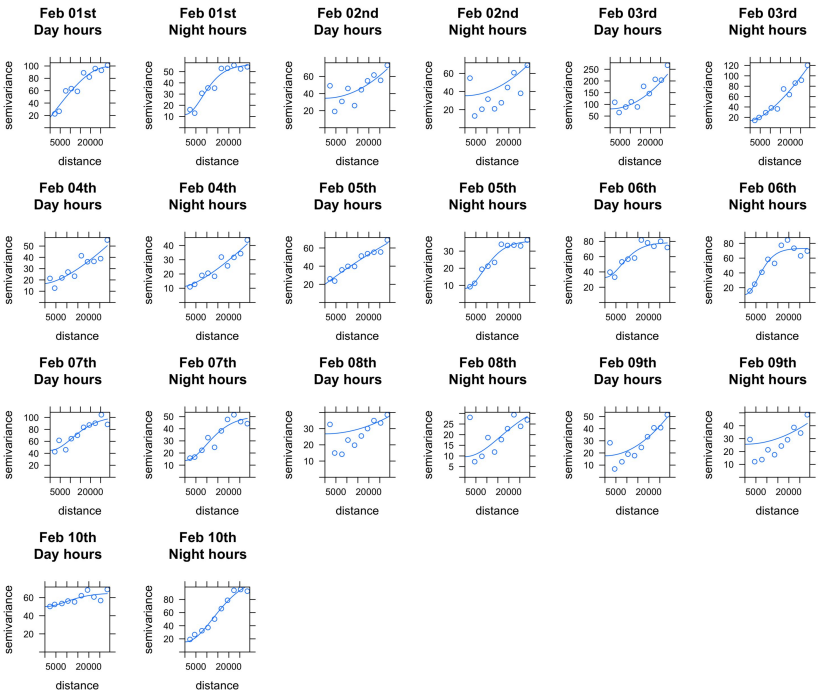


Fig. C.58 PM_{10}: Semivariogram for Kriged Maps between Feb 1st-10th

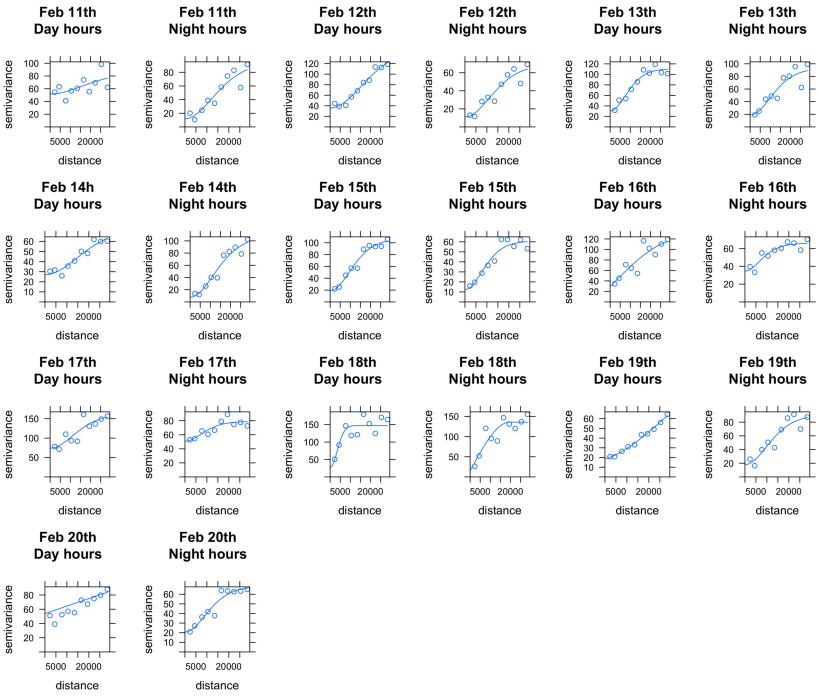


Fig. C.59 PM_{10}: Semivariogram for Kriged Maps between Feb 11th-20th

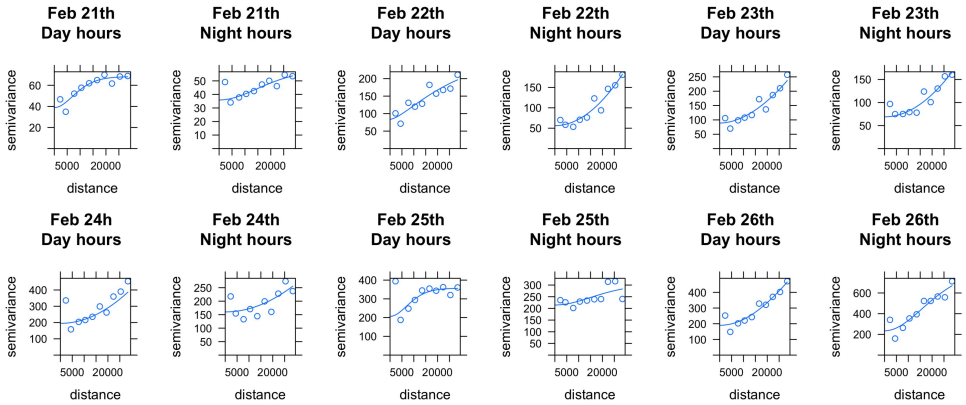


Fig. C.60 PM_{10}: Semivariogram for Kriged Maps between Feb 21th-28th

C.1.5 PM_{10}: Kriging - Predictions

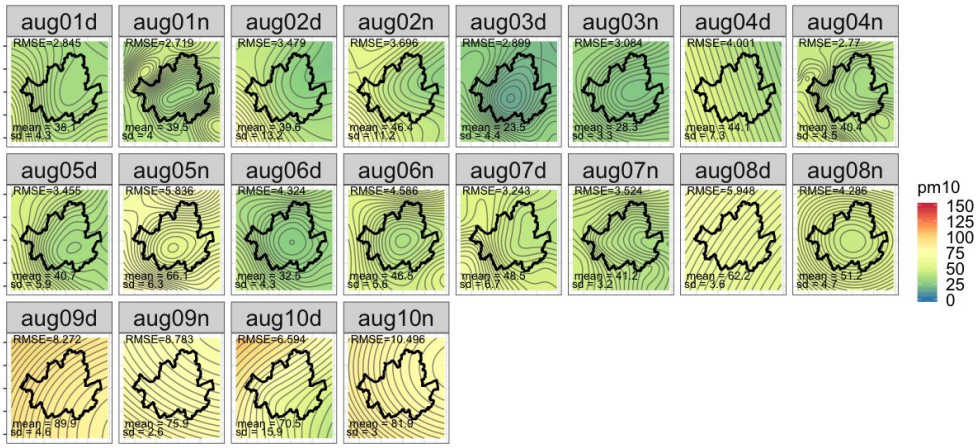


Fig. C.61 PM_{10}: Kriged Maps between Aug 1st-10th (units: $\mu g/m^3$)

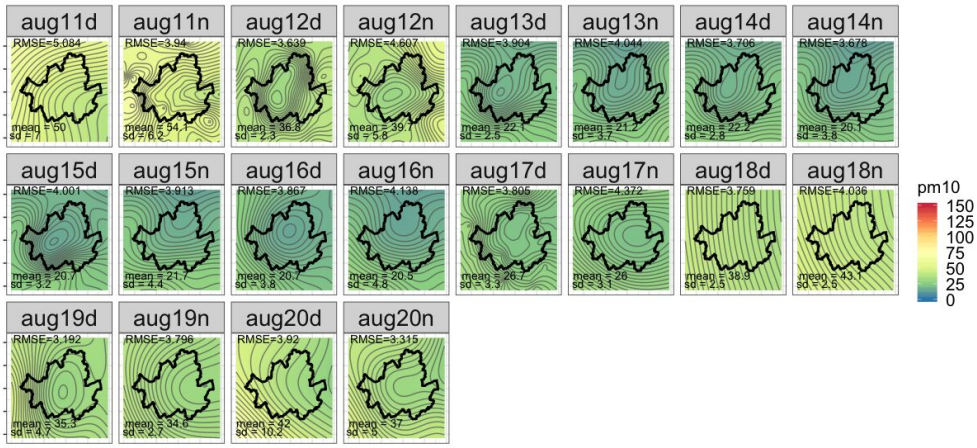


Fig. C.62 PM_{10}: Kriged Maps between Aug 11th-20th (units: $\mu g/m^3$)

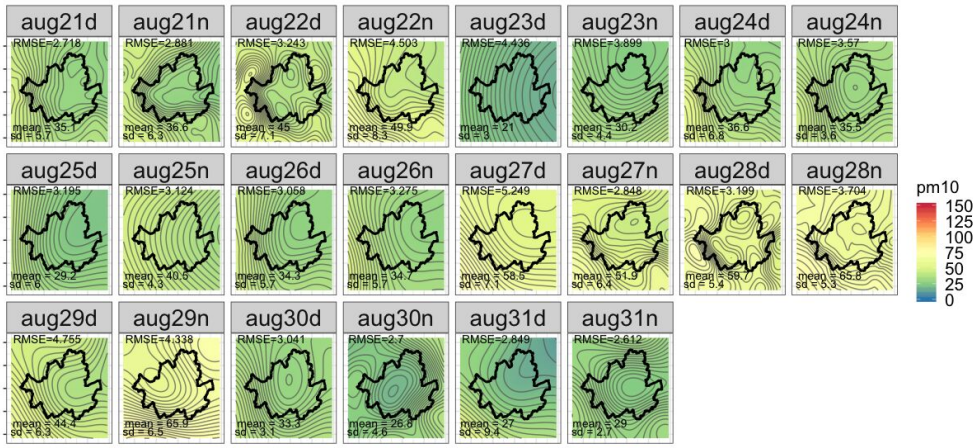


Fig. C.63 PM_{10}: Kriged Maps between Aug 21th-31st (units: $\mu g/m^3$)

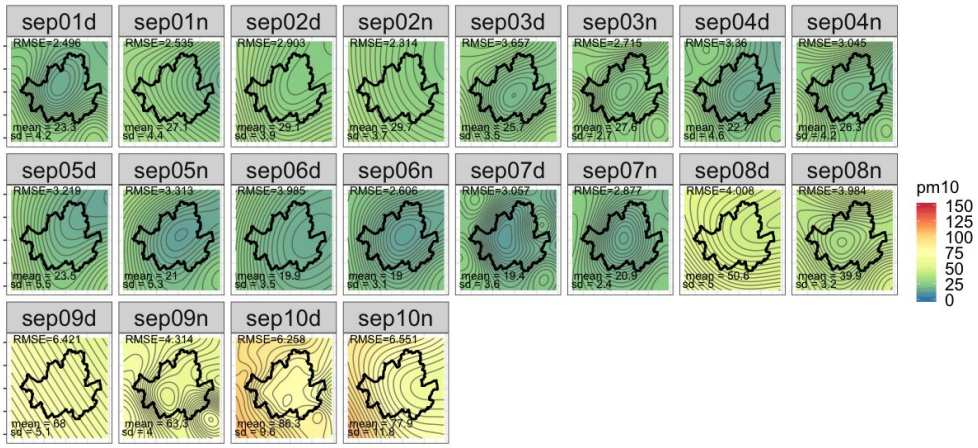


Fig. C.64 PM_{10}: Kriged Maps between Sep 1st-10th (units: $\mu g/m^3$)

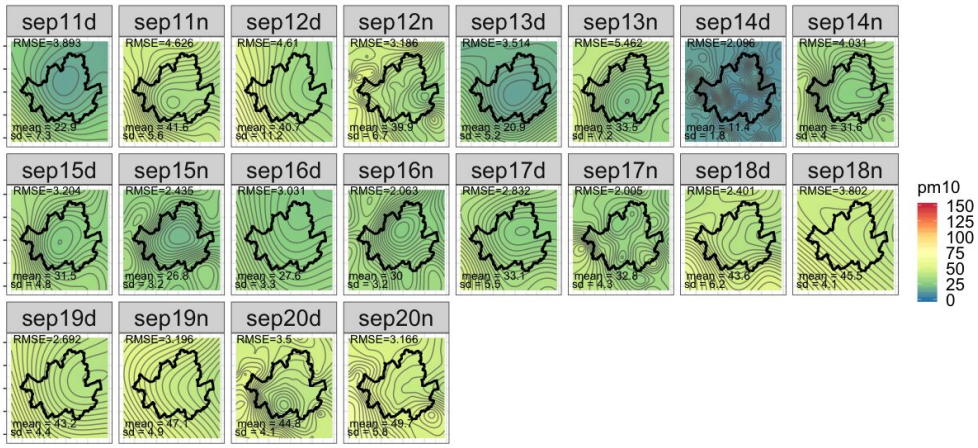


Fig. C.65 PM_{10}: Kriged Maps between Sep 11th-20th (units: $\mu g/m^3$)

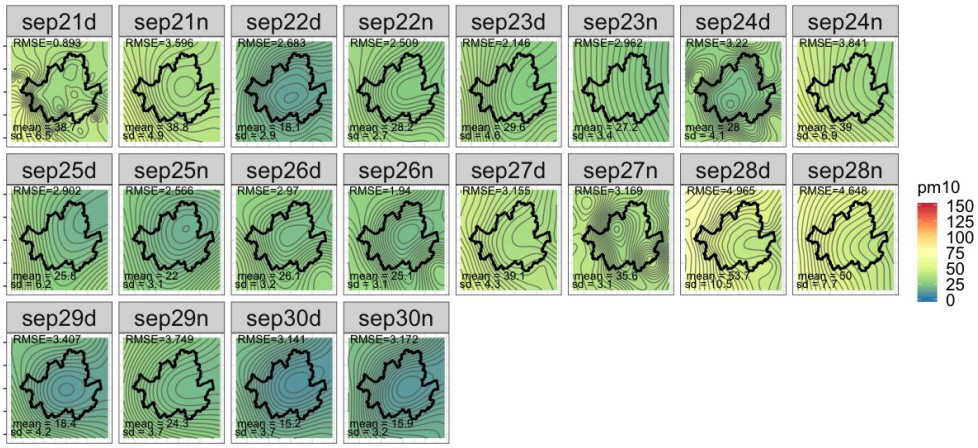


Fig. C.66 PM_{10}: Kriged Maps between Sep 21st-30th (units: $\mu g/m^3$)

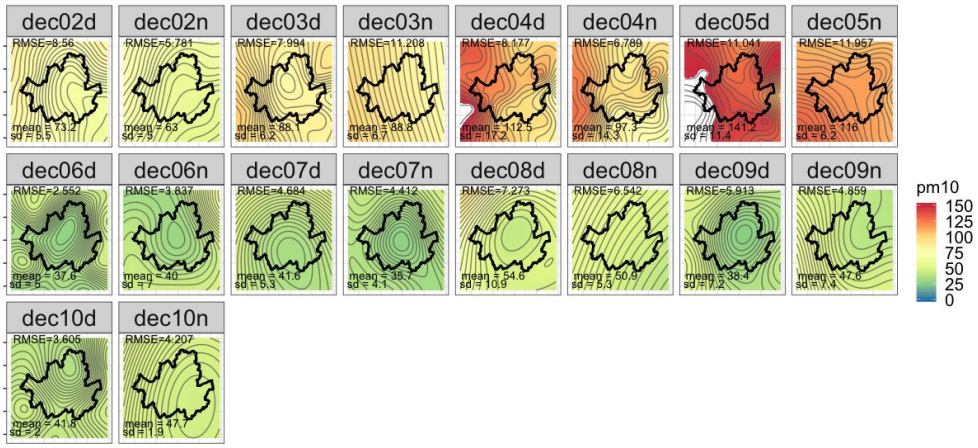


Fig. C.67 PM_{10}: Kriged Maps between Dec 1st-10th (units: µg/m^3)

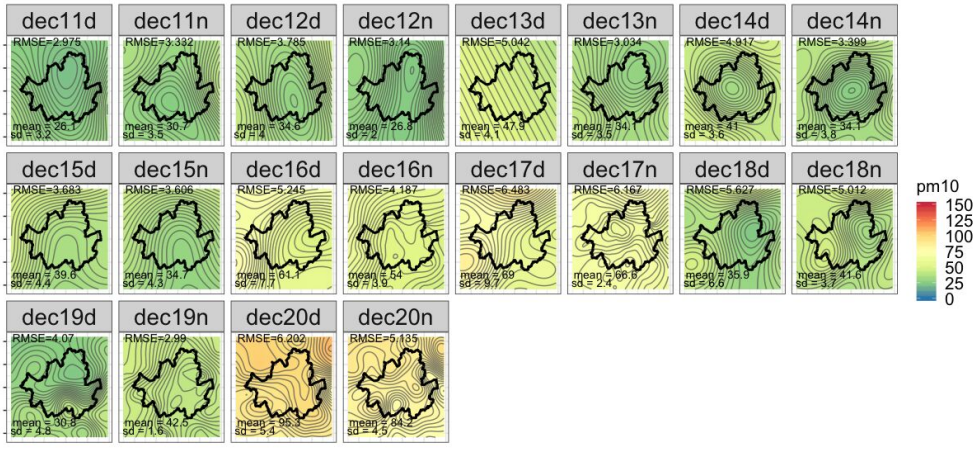


Fig. C.68 PM_{10}: Kriged Maps between Dec 11th-20th (units: µg/m^3)

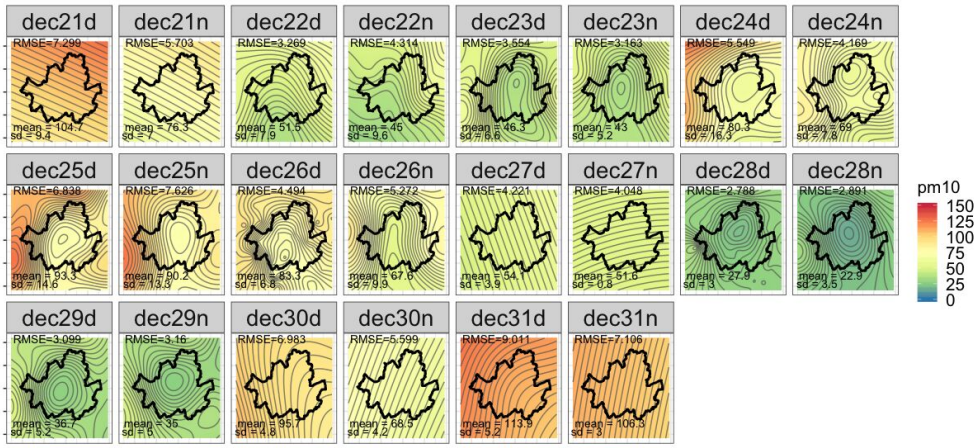


Fig. C.69 PM_{10}: Kriged Maps between Dec 21th-31st (units: µg/m^3)

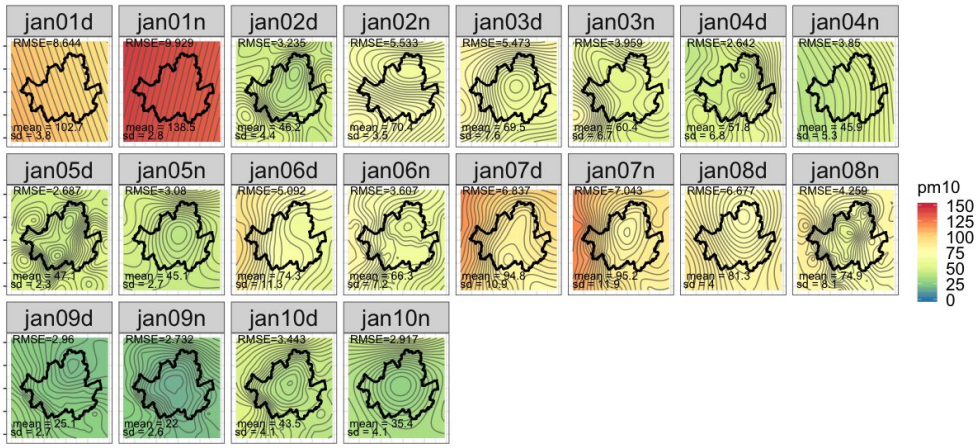


Fig. C.70 PM_{10}: Kriged Maps between Jan 1st-10th (units: µg/m^3)

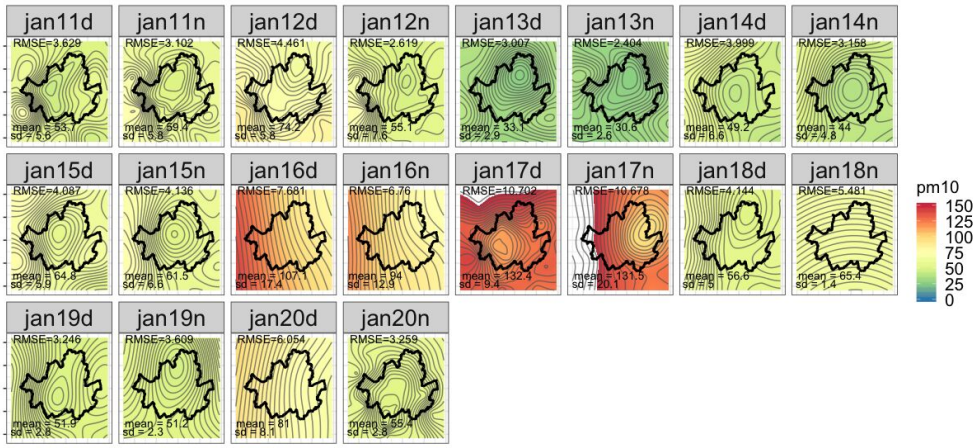


Fig. C.71 PM_{10}: Kriged Maps between Jan 11th-20th (units: $\mu g/m^3$)

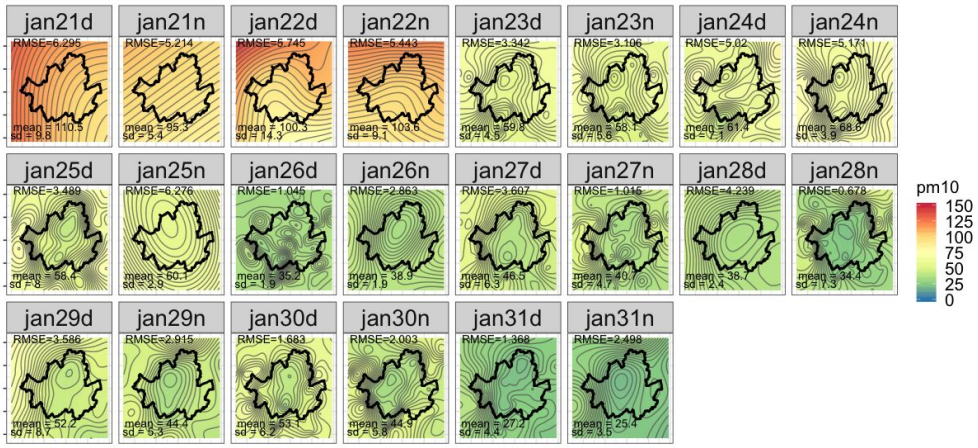


Fig. C.72 PM_{10}: Kriged Maps between Jan 21th-31st (units: $\mu g/m^3$)

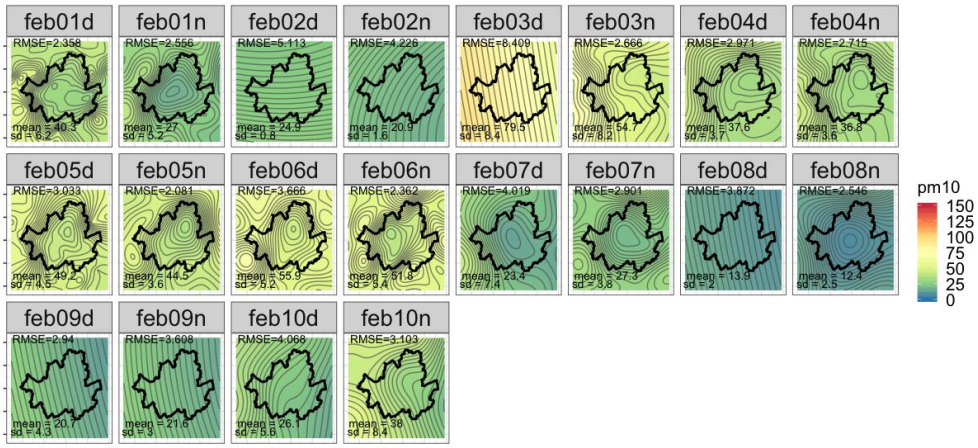


Fig. C.73 PM_{10}: Kriged Maps between Feb 1st-10th (units: µg/m^3)

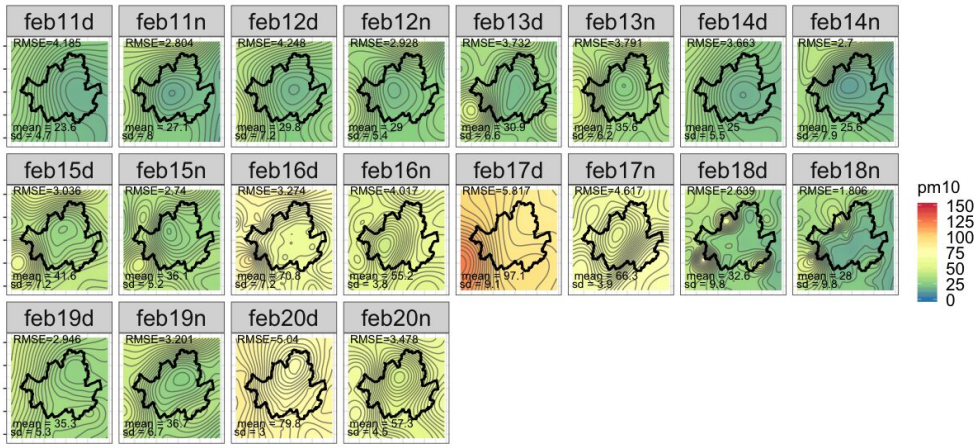


Fig. C.74 PM_{10}: Kriged Maps between Feb 11th-20th (units: µg/m^3)

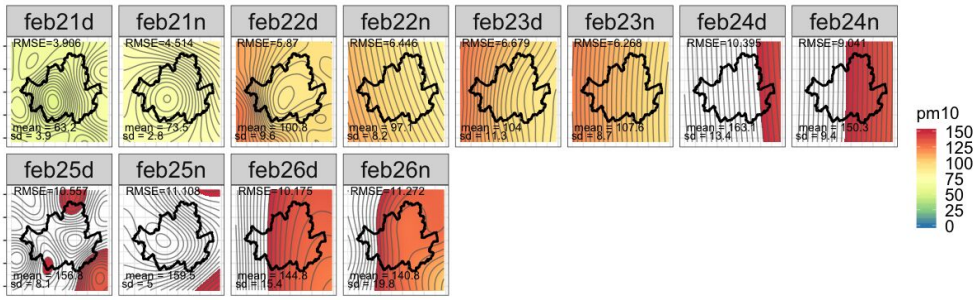


Fig. C.75 PM_{10}: Kriged Maps between Feb 21th-28th (units: µg/m^3)

C.1.6 PM_{10}: Kriging - Predictions with Addtional Road Weight

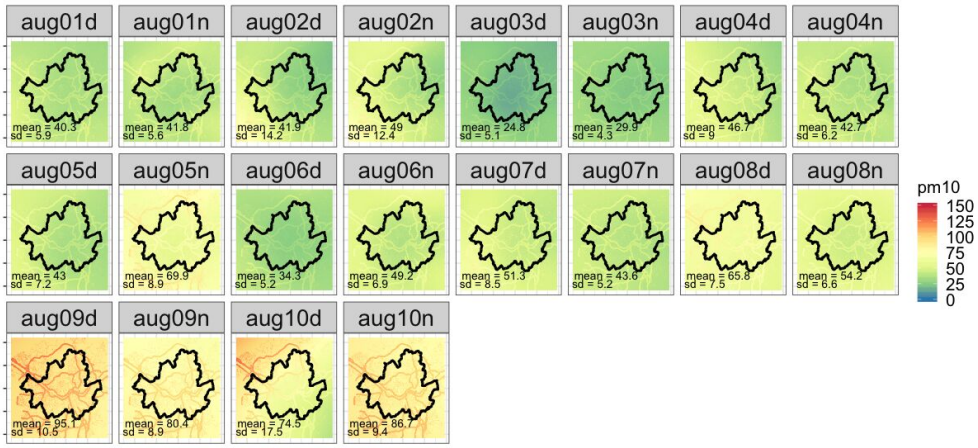


Fig. C.76 PM_{10}: Kriged Maps with Road Weights between Aug 1st-10th (units: µg/m^3)

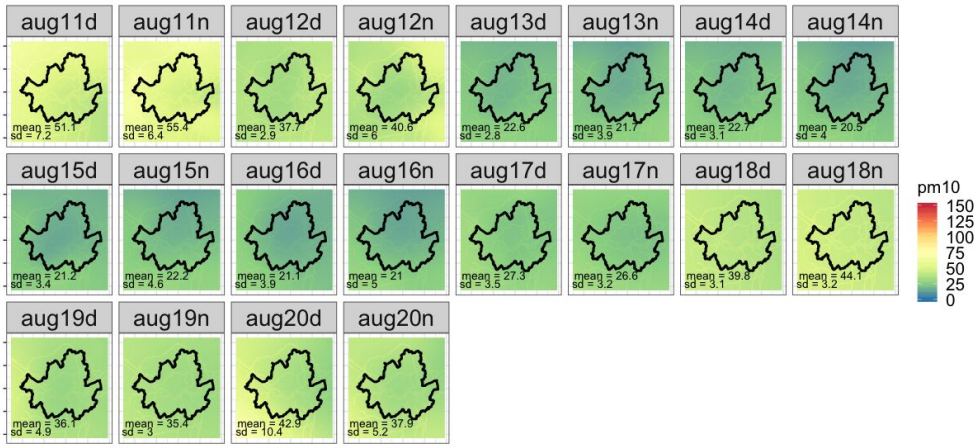


Fig. C.77 PM_{10}: Kriged Maps with Road Weights between Aug 11th-20th (units: µg/m^3)

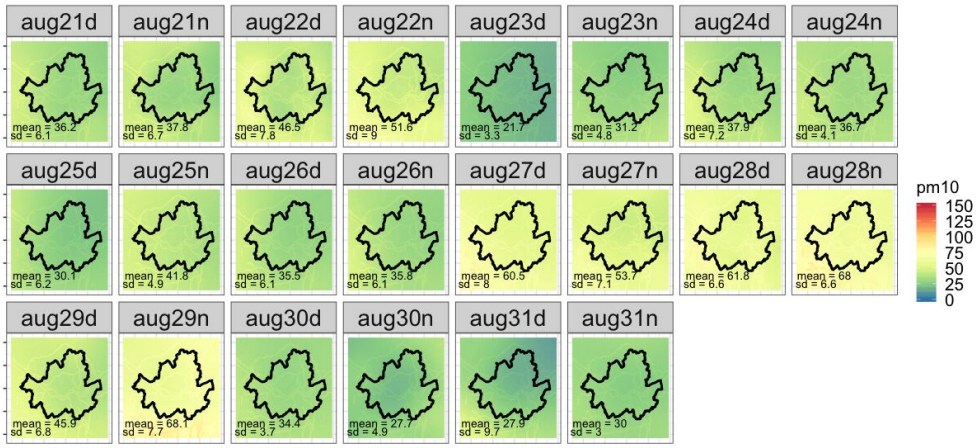


Fig. C.78 PM_{10}: Kriged Maps with Road Weights between Aug 21th-31st (units: $\mu g/m^3$)

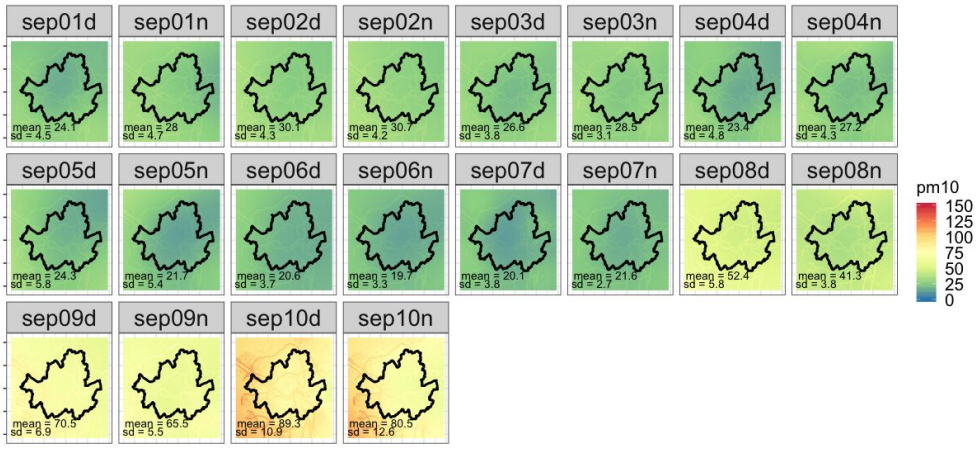


Fig. C.79 PM_{10}: Kriged Maps with Road Weights between Sep 1st-10th (units: $\mu g/m^3$)

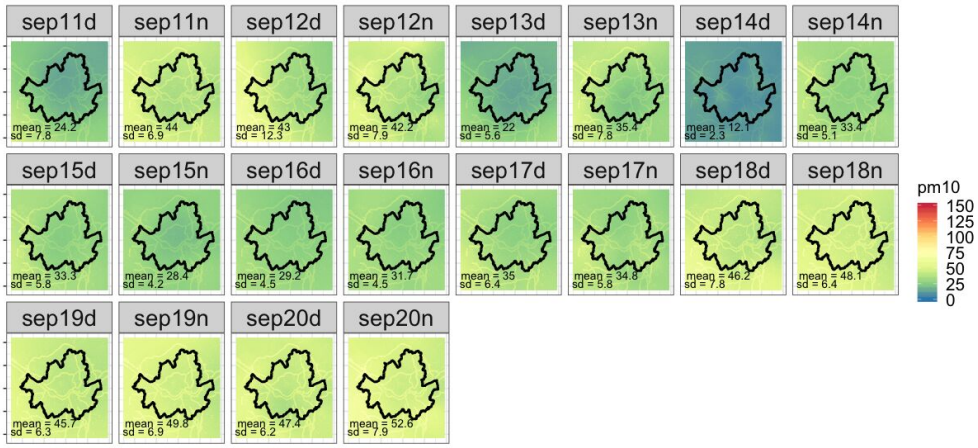


Fig. C.80 PM_{10}: Kriged Maps with Road Weights between Sep 11th-20th (units: $\mu g/m^3$)

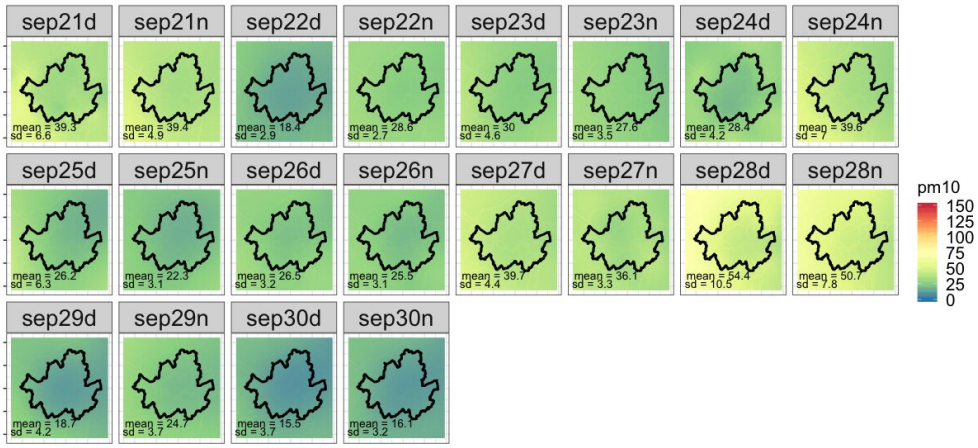


Fig. C.81 PM_{10}: Kriged Maps with Road Weights between Sep 21st-30th (units: $\mu g/m^3$)

Fig. C.82 PM_{10}: Kriged Maps with Road Weights between Dec 1st-10th (units: $\mu g/m^3$)

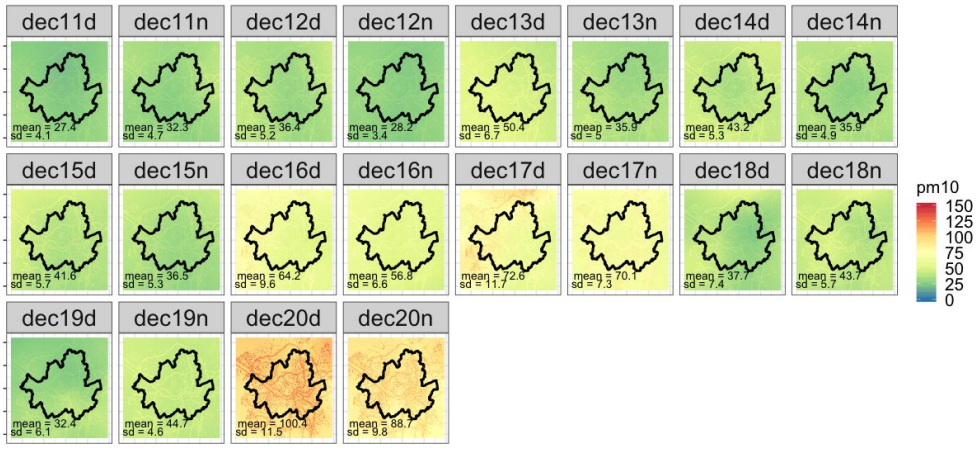


Fig. C.83 PM_{10}: Kriged Maps with Road Weights between Dec 11th-20th (units: $\mu g/m^3$)

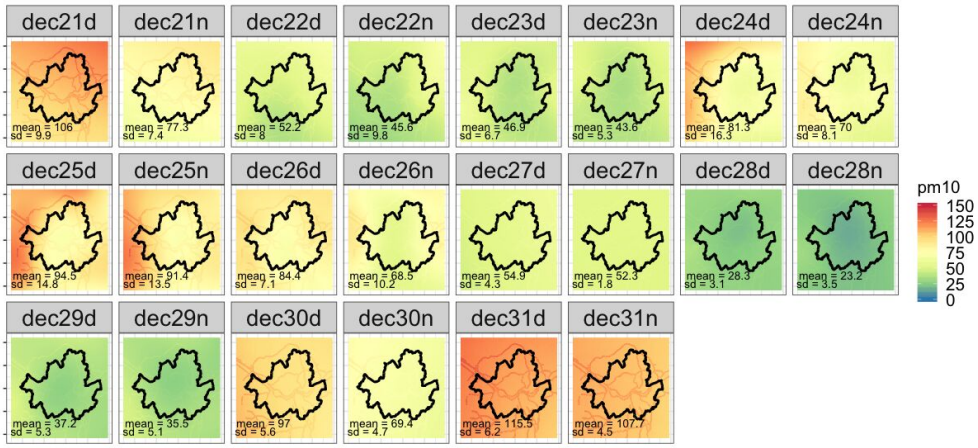


Fig. C.84 PM_{10}: Kriged Maps with Road Weights between Dec 21th-31st (units: $\mu g/m^3$)

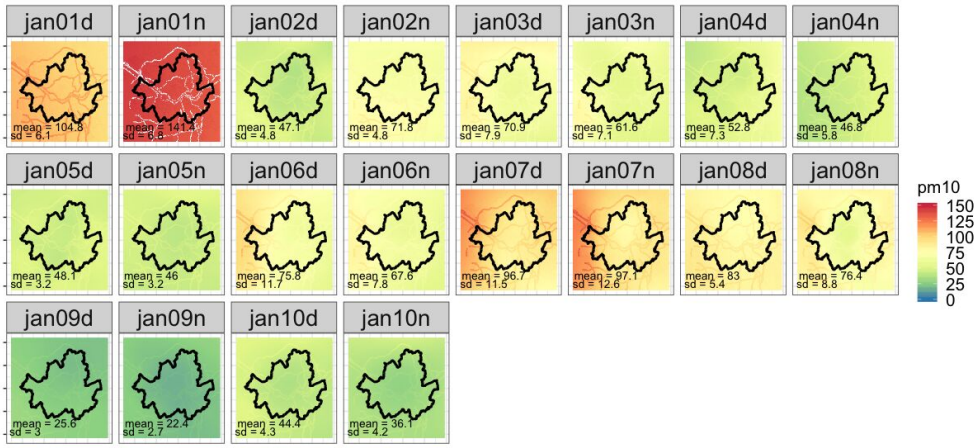


Fig. C.85 PM_{10}: Kriged Maps with Road Weights between Jan 1st-10th (units: $\mu g/m^3$)

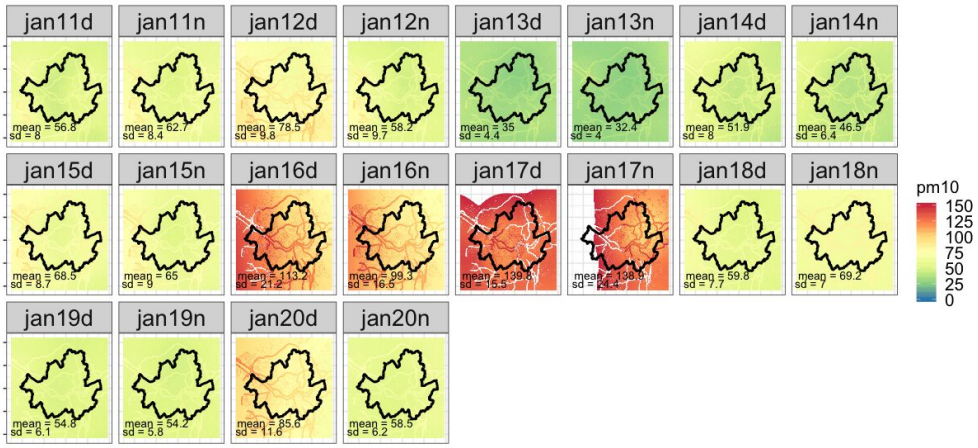


Fig. C.86 PM_{10}: Kriged Maps with Road Weights between Jan 11th-20th (units: $\mu g/m^3$)

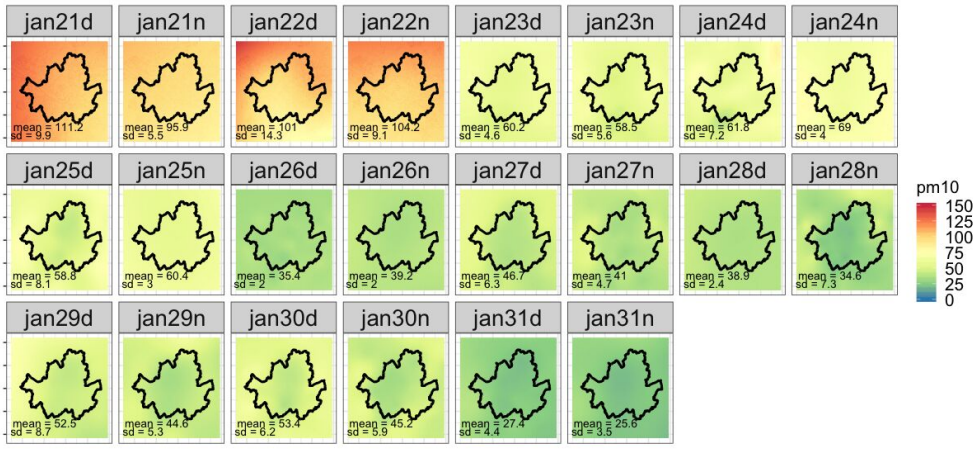


Fig. C.87 PM_{10}: Kriged Maps with Road Weights between Jan 21th-31st (units: $\mu g/m^3$)

Fig. C.88 PM_{10}: Kriged Maps with Road Weights between Feb 1st-10th (units: µg/m^3)

Fig. C.89 PM_{10}: Kriged Maps with Road Weights between Feb 11th-20th (units: µg/m^3)

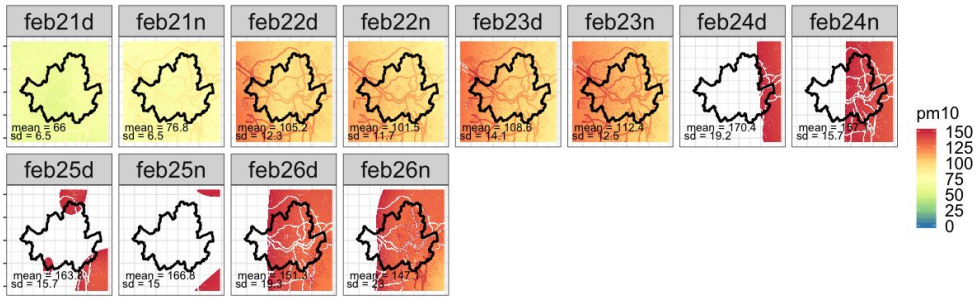


Fig. C.90 PM_{10}: Kriged Maps with Road Weights between Feb 21th-28th (units: $\mu g/m^3$)

C.2 Generalised Additive Modelling

C.2.1 NO_2: Prediction

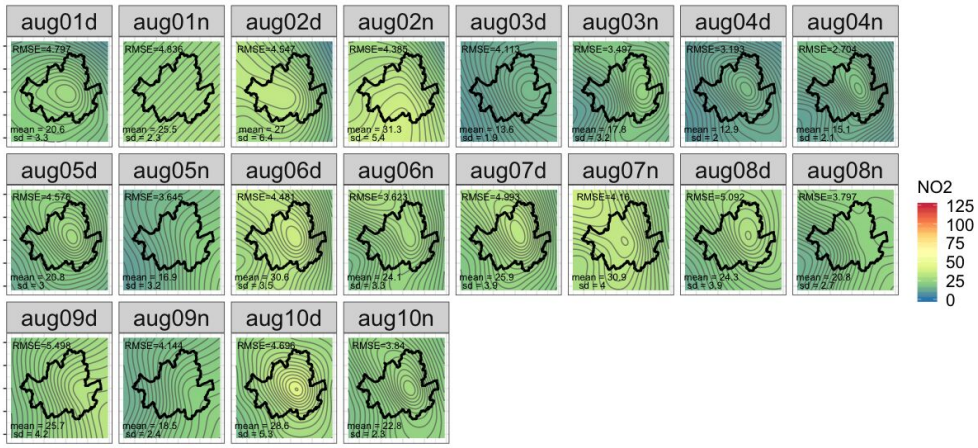


Fig. C.91 NO_2: GAM Maps between Aug 1st-10th (units: ppb)

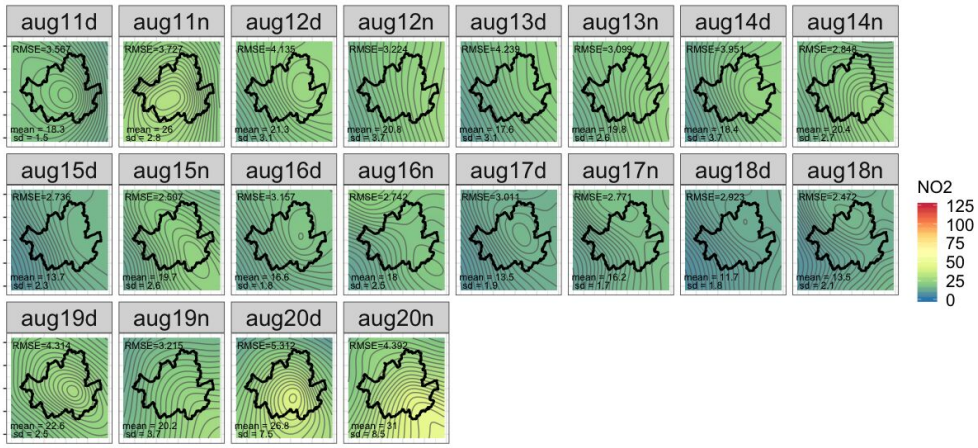


Fig. C.92 NO_2: GAM Maps between Aug 11th-20th (units: ppb)

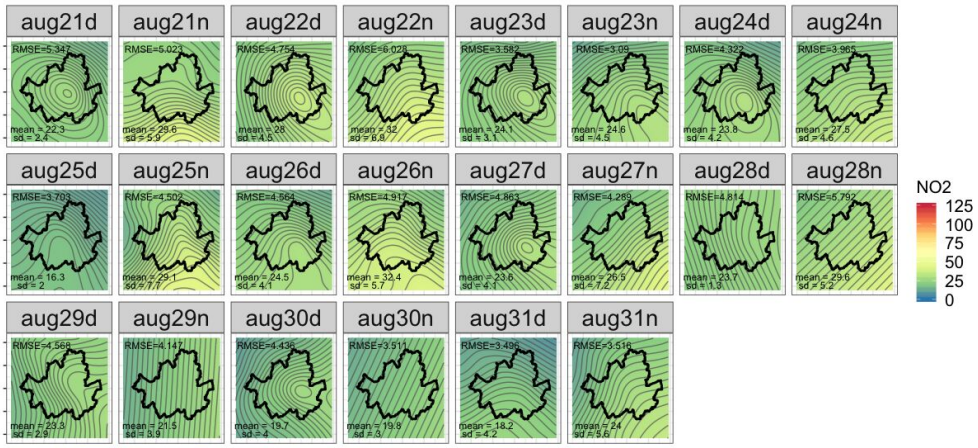


Fig. C.93 NO_2: GAM Maps between Aug 21th-31st (units: ppb)

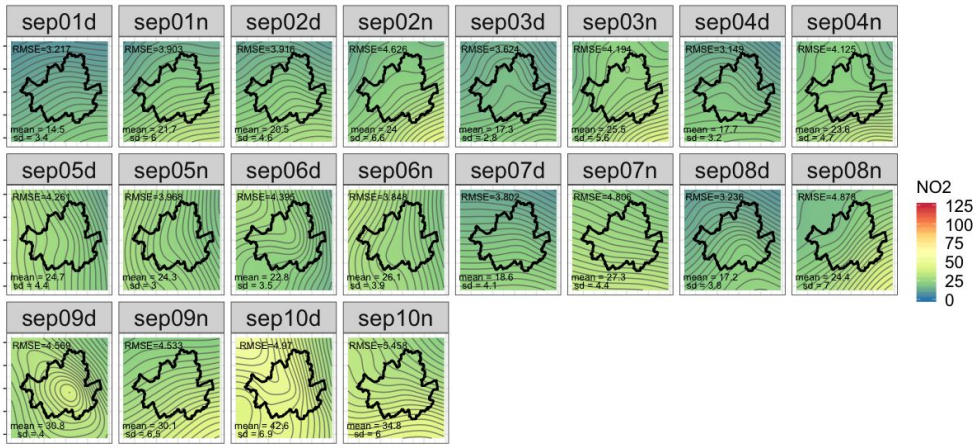


Fig. C.94 NO_2: GAM Maps between Sep 1st-10th (units: ppb)

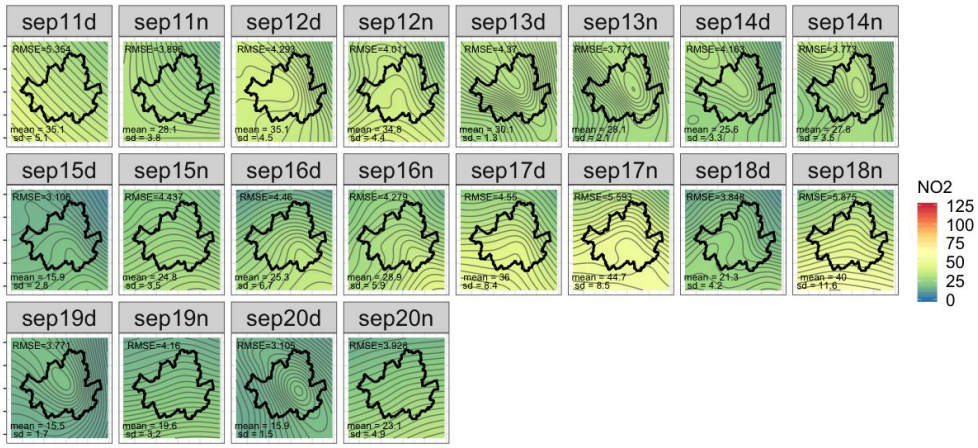


Fig. C.95 NO_2: GAM Maps between Sep 11th-20th (units: ppb)

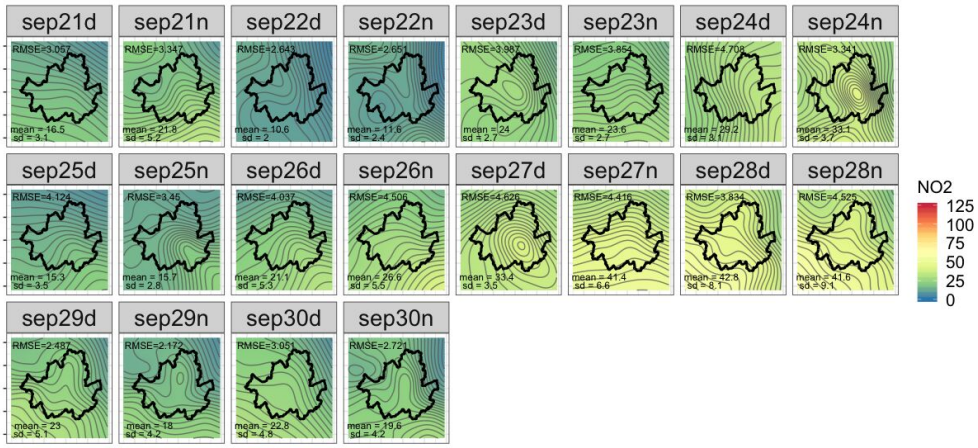


Fig. C.96 NO_2: GAM Maps between Sep 21st-30th (units: ppb)

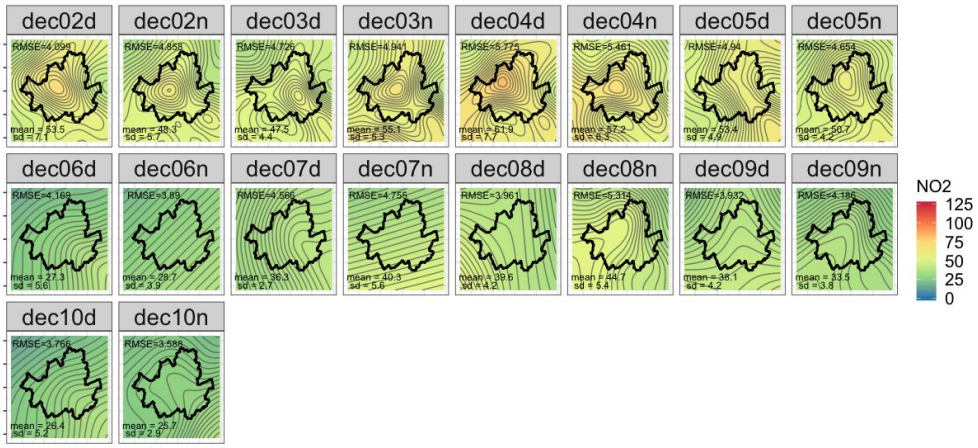


Fig. C.97 NO_2: GAM Maps between Dec 1st-10th (units: ppb)

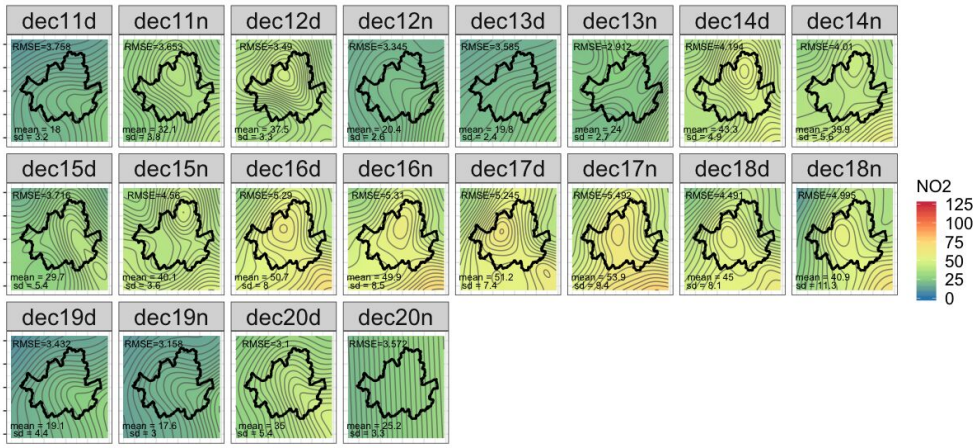


Fig. C.98 NO_2: GAM Maps between Dec 11th-20th (units: ppb)

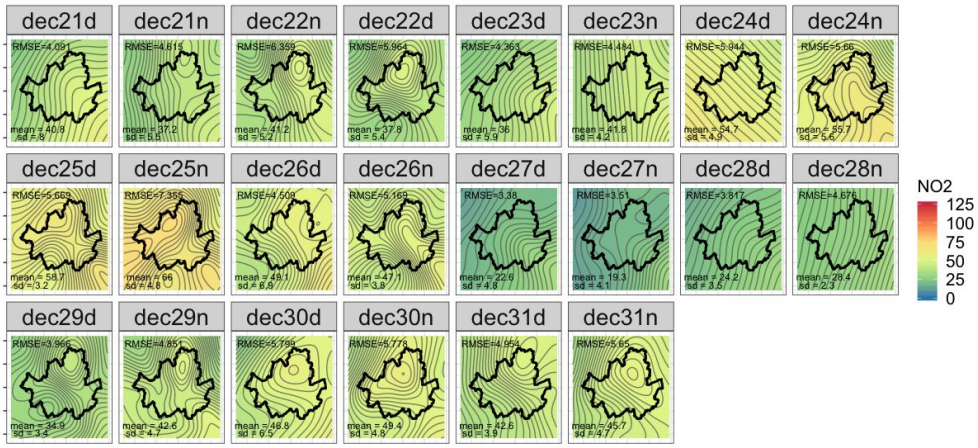


Fig. C.99 NO_2: GAM Maps between Dec 21th-31st (units: ppb)

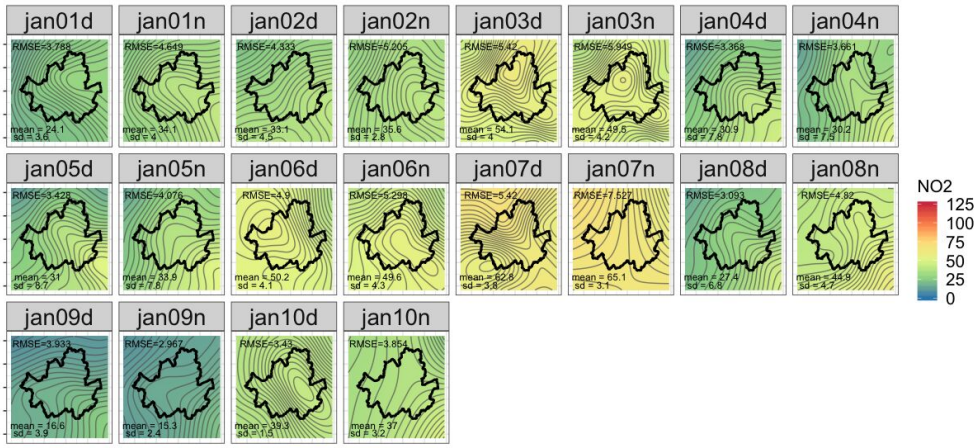


Fig. C.100 NO_2: GAM Maps between Jan 1st-10th (units: ppb)

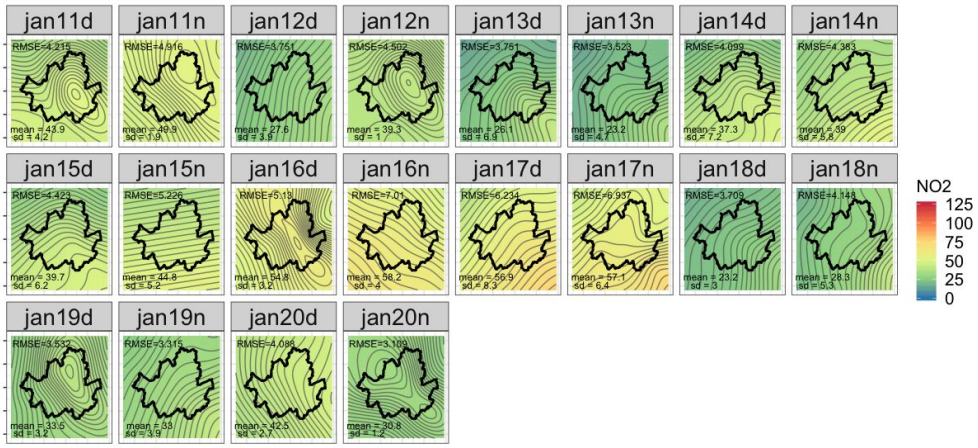


Fig. C.101 NO_2: GAM Maps between Jan 11th-20th (units: ppb)

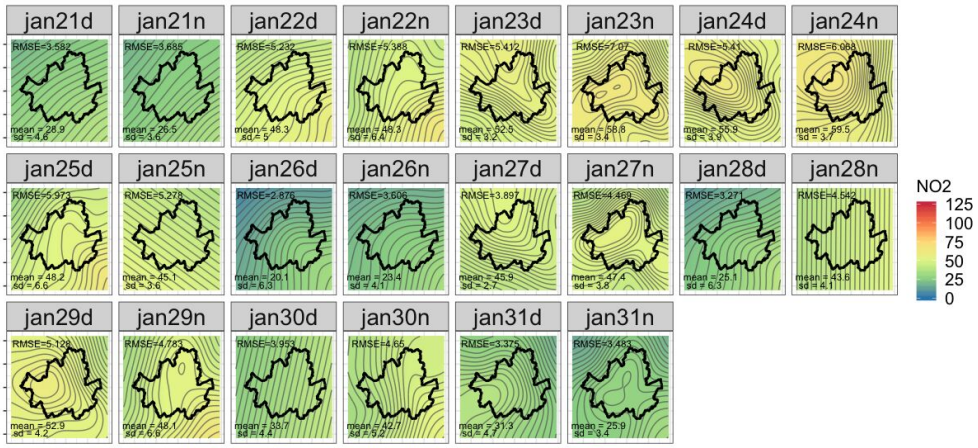


Fig. C.102 NO_2: GAM Maps between Jan 21th-31st (units: ppb)

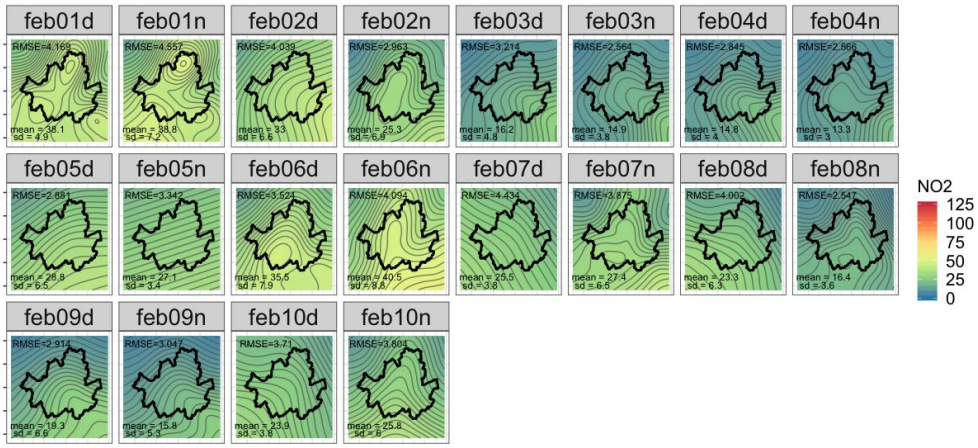


Fig. C.103 NO_2: GAM Maps between Feb 1st-10th (units: ppb)

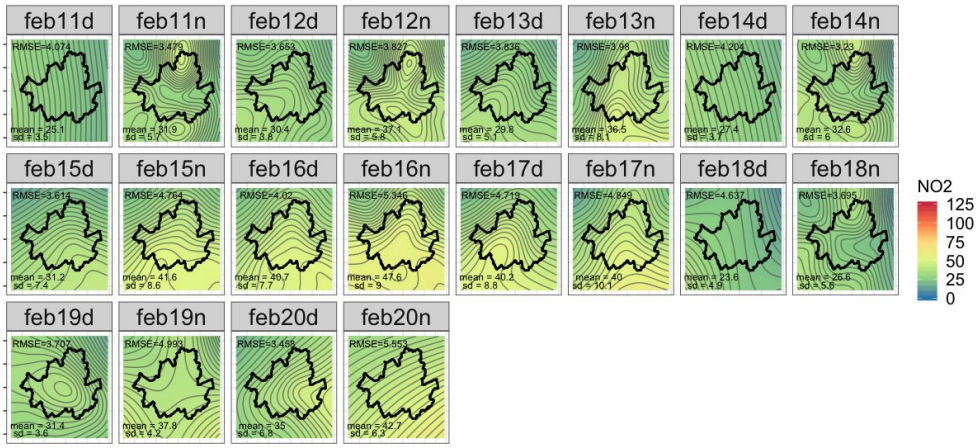


Fig. C.104 NO_2: GAM Maps between Feb 11th-20th (units: ppb)

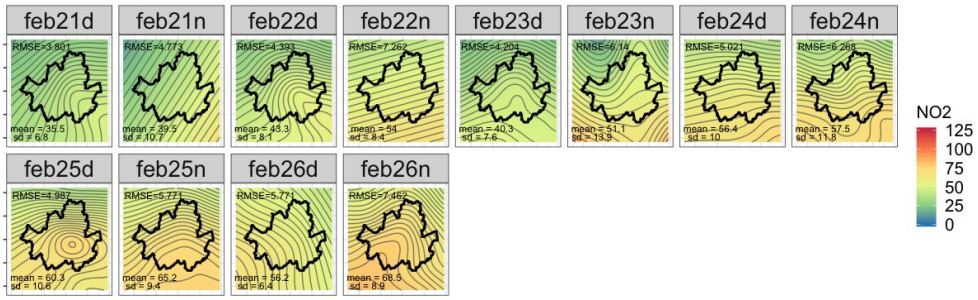


Fig. C.105 NO_2: GAM Maps between Feb 21th-28th (units: ppb)

C.2.2 NO_2: Prediction

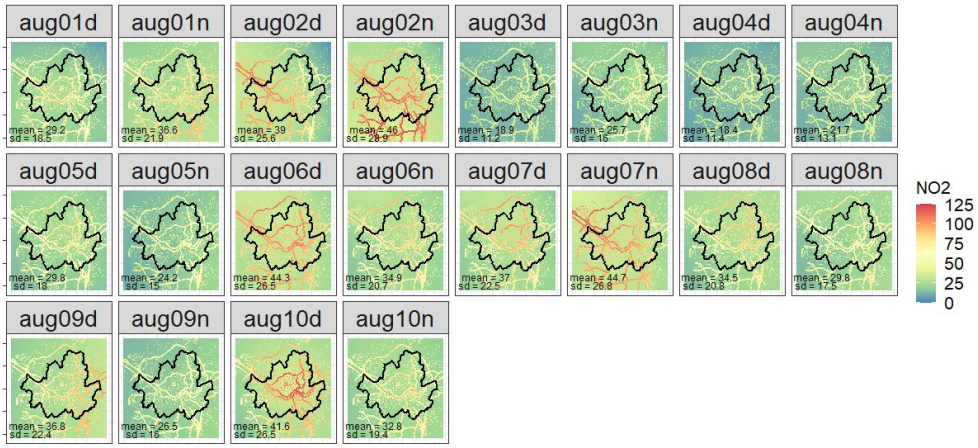


Fig. C.106 NO_2: GAM Maps between Aug 1st-10th (units: ppb)

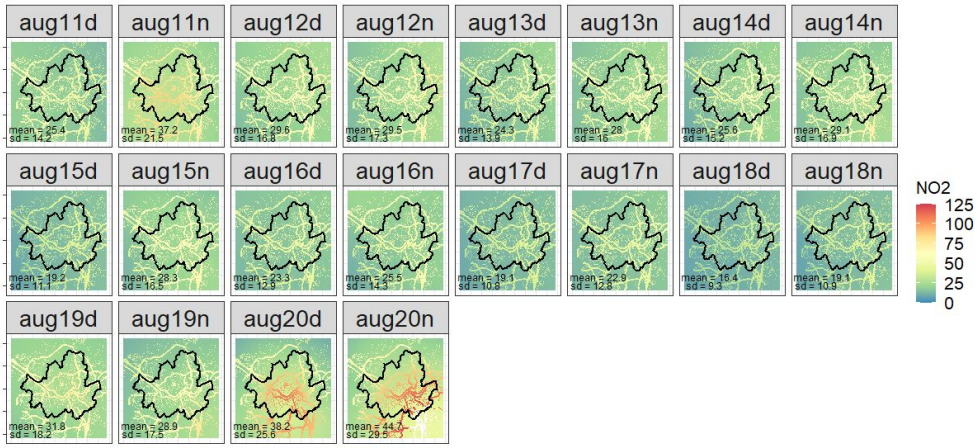


Fig. C.107 NO_2: GAM Maps between Aug 11th-20th (units: ppb)

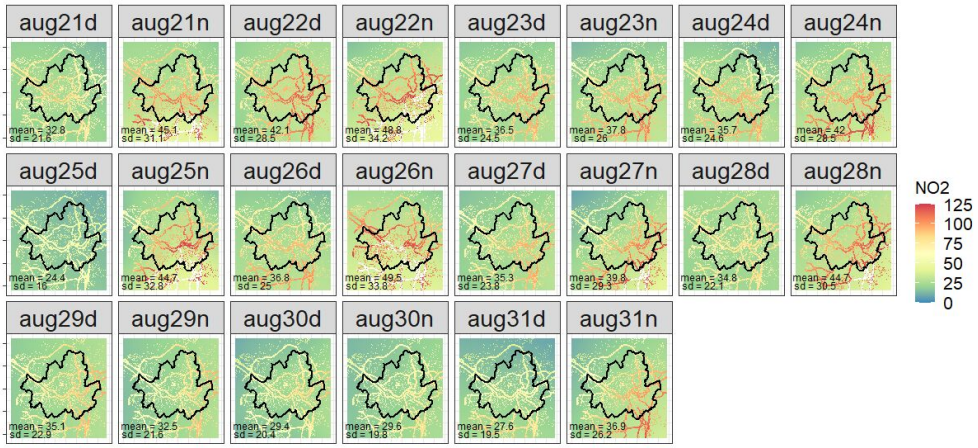


Fig. C.108 NO_2: GAM Maps between Aug 21th-31st (units: ppb)

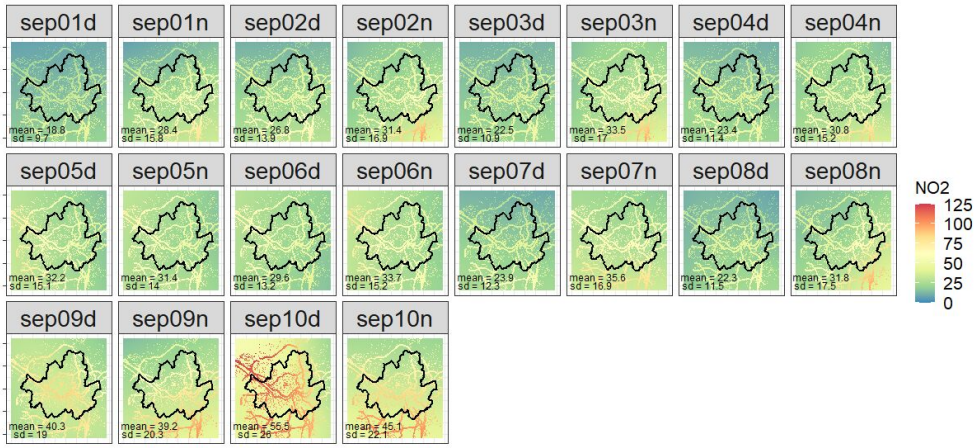


Fig. C.109 NO_2: GAM Maps between Sep 1st-10th (units: ppb)

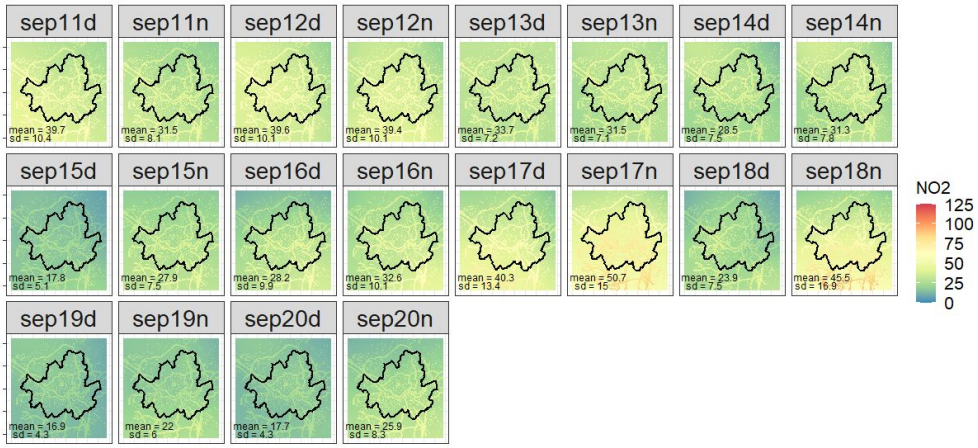


Fig. C.110 NO_2: GAM Maps between Sep 11th-20th (units: ppb)

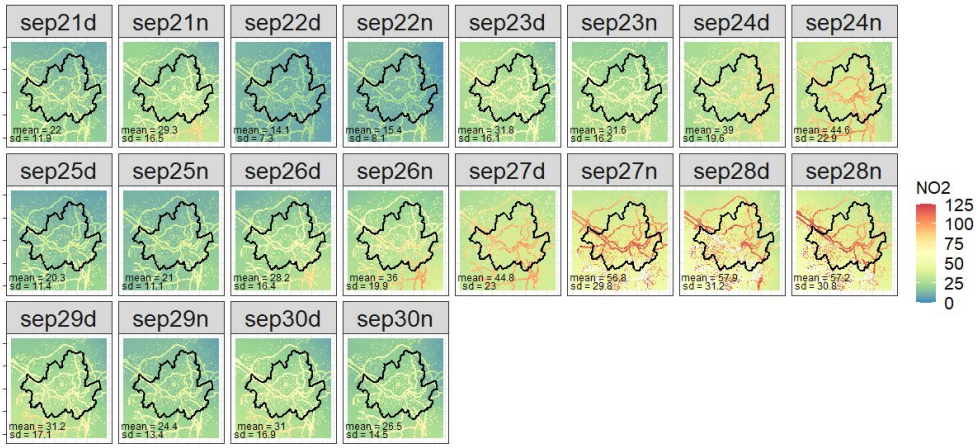


Fig. C.111 NO_2: GAM Maps between Sep 21st-30th (units: ppb)

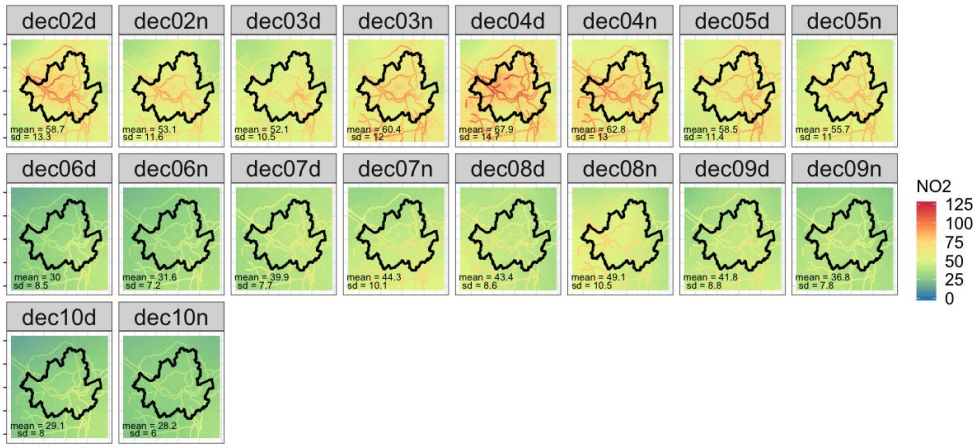


Fig. C.112 NO_2: GAM Maps between Dec 1st-10th (units: ppb)

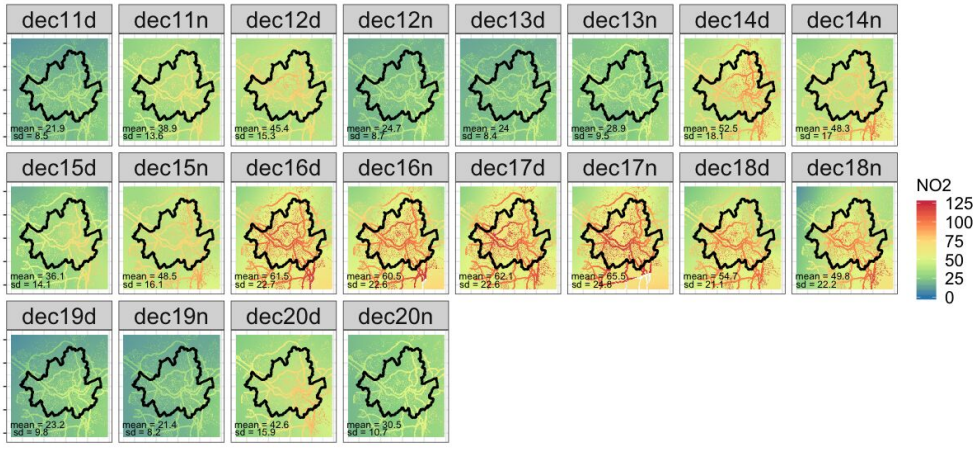


Fig. C.113 NO_2: GAM Maps between Dec 11th-20th (units: ppb)

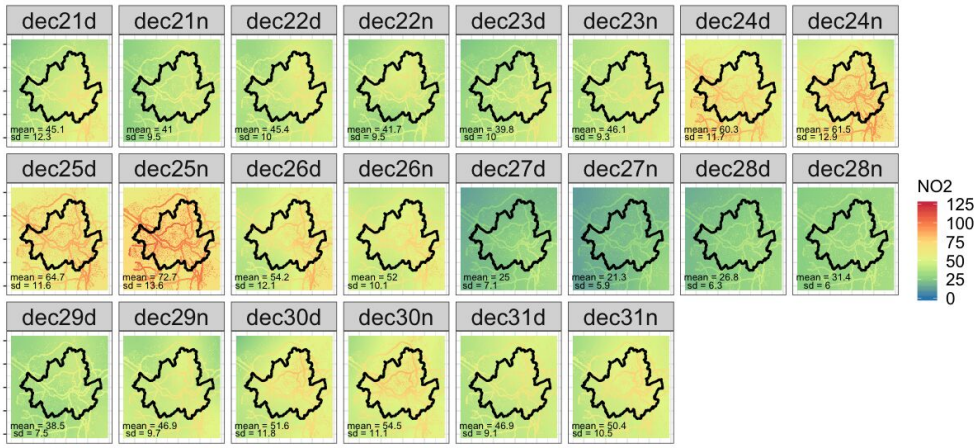


Fig. C.114 NO_2: GAM Maps between Dec 21th-31st (units: ppb)

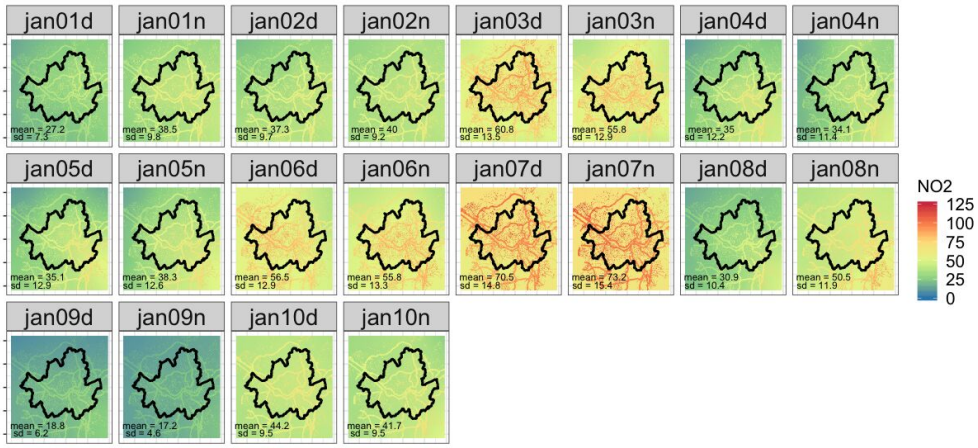


Fig. C.115 NO_2: GAM Maps between Jan 1st-10th (units: ppb)

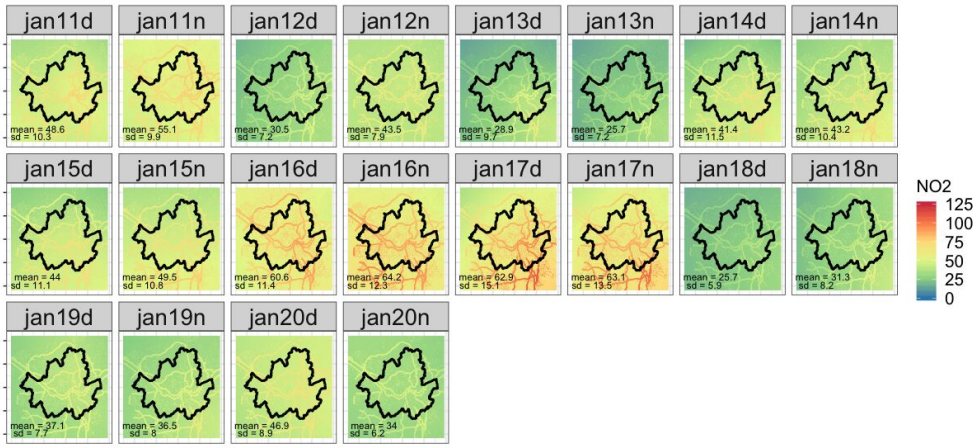


Fig. C.116 NO_2: GAM Maps between Jan 11th-20th (units: ppb)

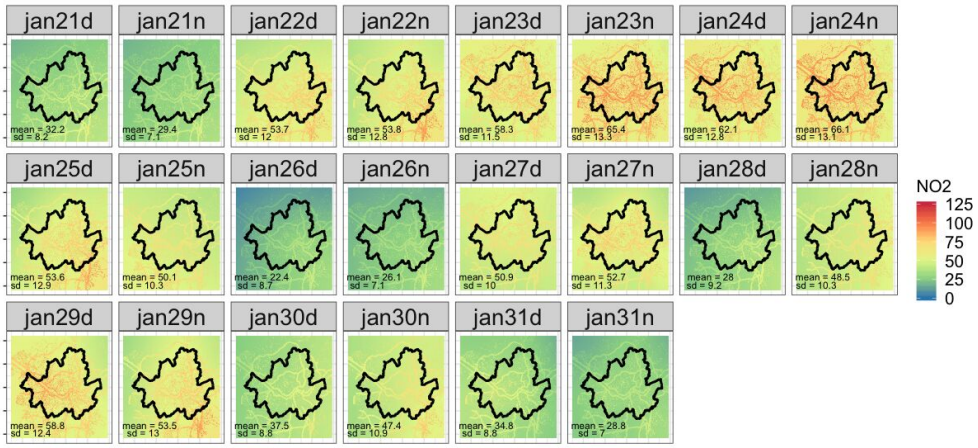


Fig. C.117 NO_2: GAM Maps between Jan 21th-31st (units: ppb)

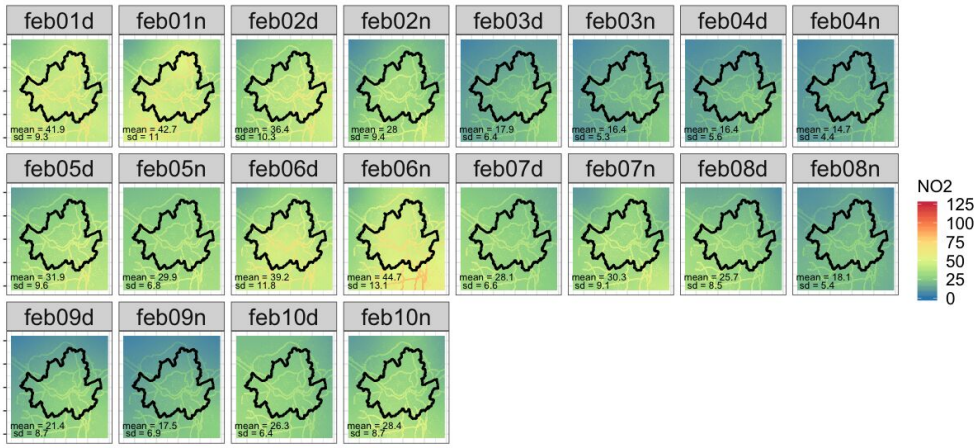


Fig. C.118 NO_2: GAM Maps between Feb 1st-10th (units: ppb)

Fig. C.119 NO_2: GAM Maps between Feb 11th-20th (units: ppb)

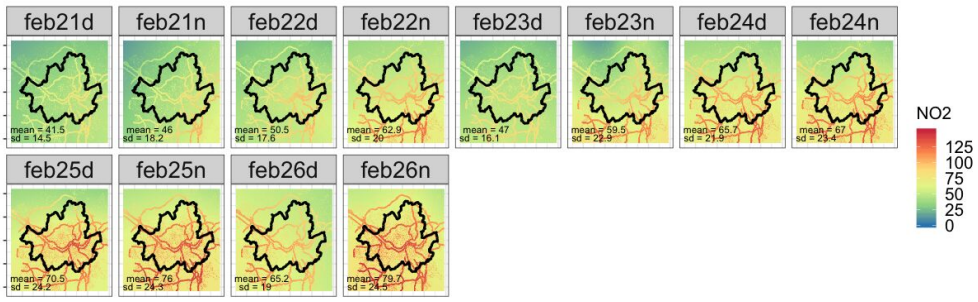


Fig. C.120 NO_2: GAM Maps between Feb 21th-28th (units: ppb)